12-2-2019

Advanced Forecasting and Optimization Techniques in Power System Planning and Operations

Xiaorong Sun
University of Connecticut - Storrs, xiaorong.sun@uconn.edu

Advanced Forecasting and Optimization Techniques in Power System Planning and Operations

Xiaorong Sun, PhD

University of Connecticut, 2019

Utility companies, Independent System Operators, and Transmission Operators require sophisticated techniques in short-term load forecasting, unit commitment (UC), and dynamic line rating in view of the complicated customer behavior, increasing integration of renewable energy, and characteristics of transmission lines in power systems. In this thesis, a set of novel methodologies are developed to address the challenges posed by each problem.

1. Short-term load forecasting at the distribution level is difficult in view of the complicated load features, the large number of distribution-level nodes, and possible switching operations. A hierarchical forecasting approach is established to capture load characteristics at different levels. Load of a root node at any user-defined subtree is first forecast by a wavelet neural network with appropriate inputs. Child nodes categorized as “regular” and “irregular” based on load pattern similarities are forecast separately. Switching operation detection and follow-up adjustments are performed to capture abnormal changes and improve forecasting accuracy.

2. With a large number of combined cycle (CC) units represented by configuration-based modeling, solving UC problem through the state-of-the-practice branch-and-cut method suffers from poor performance. The recently developed Surrogate Lagrangian Relaxation is significantly enhanced through adding quadratic penalties on constraint violations to accelerate convergence. Quadratic penalty terms are linearized through a novel use of absolute value functions. Therefore, resource-level subproblems can be formulated and solved by branch-and-cut. Complicated constraints within a CC are thus handled within a subproblem. Subproblem solutions are then effectively coordinated.

3. Employing dynamic thermal rating, which adapts the thermal capacity of an overhead transmission line based on dynamic weather, is difficult in view of many weather factors, weather data availability, weather uncertainties, and the topology of transmission lines. In the developed probabilistic forecasting model, major weather factors are selected based on impact analysis and are modeled through a spatio-temporal regression with all available data sources. Spatial topology and weather uncertainties are simultaneously captured by treating the line rating as the minimum of critical span thermal capacities. An approximation is applied to determine the distribution and extract appropriate percentiles with light computational costs.

Advanced Forecasting and Optimization Techniques in Power System Planning and Operations

Xiaorong Sun

B.S., Hohai University, Nanjing, China, 2010

M.S., University of Connecticut, Storrs, CT, 2014

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2019

APPROVAL PAGE

Doctor of Philosophy Dissertation

Advanced Forecasting and Optimization Techniques in Power System Planning and Operations

Presented by

Xiaorong Sun, B.S., M.S.

Major Advisor _____

Peter B. Luh

Associate Advisor _____

Yaakov Bar-Shalom

Associate Advisor _____

Laurent Michel

University of Connecticut
2019

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Peter. B. Luh for the continuous support of my doctoral study and related research, for his guidance, encouragement, patience, motivation, enthusiasm, and immense knowledge helped me in all the time of research and paved the way for my successful dissertation.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Yaakov Bar-Shalom and Prof. Laurent Michel, for their insightful comments and encouragement, but also for the hard questions which incented to widen my research from various perspectives.

I gratefully acknowledge Dr. Kwok Cheung, Dr. Jie Wan, and Dr. Wei Guan, Dr. Yonghong Chen, who provided me opportunities to join their team as an intern, and gave access to the utility and ISOs' data. Without their previous support, it would not be possible to conduct this research.

I would like to thank my fellow labmates, Che Guan, Congcong Wang, Mikhail Bragian, Bing Yan, Yaowen Yu, Ying Yan, Danxu Zhang and Xuesong Lu in the Manufacturing Systems Laboratory group and my classmates for the stimulating discussions, pleasant studies and enjoyable hours, which enriched my experience.

And last but not the least, I would like to thank my family: my parents and my younger brother for supporting me spiritually throughout my PhD study and my life. Great thanks go to my husband, Dr. Chenhao Jin and my two lovely sons Charles Jin and Henry Jin. I appreciate the grateful time and the power inspired by the family. Thank you for all your support.

Table of Contents

Publications Related to this Thesis

Journal Articles

[1] **X. Sun**, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. S. Venkata, and M. T. Miller, An Efficient Approach to Short-Term Load Forecasting at the Distribution Level, *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2526-2537, 2016.

[2] **X. Sun**, P. B. Luh, M. A. Bragin, Y. Chen, J. Wan, and F. Wang, A Novel Decomposition and Coordination Approach for Large-Scale Day-Ahead Unit Commitment Problems with Combined Cycle Units, *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5297-5308, Sept. 2018.

[3] M. A. Bragin, P. B. Luh, B. Yan, and **X. Sun**, A Scalable Solution Methodology for Mixed-Integer Linear Programming Problems Arising in Automation, in *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 531-541, April 2019.

[4] **X. Sun**, P. B. Luh, K. W. Cheung, and W. Guan, Probabilistic Forecasting of Dynamic Thermal Rating for Overhead Transmission Lines, *IEEE Transactions on Power Systems*, submitted, 2019.

Conference Proceedings

[5] **X. Sun**, P. B. Luh, L. D. Michel, K. W. Cheung, W. Guan, K. Chung and S. Corbo, An Efficient Approach for Short Term Substation Load Forecasting, in *Proceedings of the 2013 IEEE Power and Energy Society General Meeting* (**one of the *Best Papers on System Operations and Market Economics*),** Vancouver, BC, Canada, July 2013.

[6] W. Guan, K. Chung, K. W. Cheung, **X. Sun**, P. B. Luh, L. D. Michel, and S. Corbo, Advanced Load Forecasting with Hierarchical Forecasting Capability, in *Proc. of the 2013 IEEE Power and Energy Soc. General Meeting*, Vancouver, 2013.

[7] **X. Sun**, P. B. Luh, K. W. Cheung, and W. Guan, Probabilistic Forecasting of Dynamic Line Rating for Over-head Transmission Lines, in *Proceedings of the 2015 IEEE Power and Energy Society General Meeting*, July 2015.

[8] **X. Sun**, P. B. Luh, K. W. Cheung, and W. Guan, Forecasting Real-time Actual Net Interchange of Electric Power, in *Proceedings of the 2014 IEEE Power and Energy Society General Meeting*, National Harbor, Maryland, USA, July 2014.

[9] **X. Sun**, P. B. Luh, M. A. Bragin, Y. Chen, J. Wan, and F. Wang, A decomposition and coordination approach for large-scale security constrained unit commitment problems with combined cycle units, in *Proc. 2017 IEEE Power and Energy Soc. General Meeting* (**one of the *Best Papers on System Operations and Market Economics***), Chicago, 2017.

[10] **X. Sun**, P. B. Luh, K. W. Cheung and W. Guan, Impacts of solar penetration on short-term net load forecasting at the distribution level, submitted to *2020 IEEE Power and Energy Society General Meeting.*

Chapter 1

Introduction

1.1 Motivations

Utility companies and Independent System Operators (ISOs) require sophisticated load forecasting techniques at the distribution level in view of the complicated customer behaver. High quality load forecasting is important for the planning and operations of distribution systems [1], [2]. For instance, substation and feeder forecasts provide utilities with advanced warnings on potential substation and feeder overloading. Customer load forecasting helps utilities schedule and dispatch community storage batteries to shave peak load in the smart grid environment.

Day-ahead unit Commitment (UC) is an important problem faced by ISOs. The UC problem is formulated as a Mixed Integer Linear Programming (MILP) problem, and has specified solving time limits and solution quality requirements [3]. Considering the extended network size of ISOs and the increasing number of generation resources, solving the UC problem becomes challenging. With a large number of combined cycle units (CCs) represented by configuration-based modeling [4], solving day-ahead unit commitment (UC) problem becomes even difficult.

Transfer capability of an overhead transmission line is limited by its thermal rating. Static thermal rating based on the worst-case weather condition does not utilize dynamic cooling and heating affects to conductors from ambient weather. Dynamic thermal rating (DTR) adapts the thermal capacity based on measured and predicted weather and typically results in a high rating without scarifying system security. The DTR could be integrated with day-ahead unit commitment and power flow analysis to improve the efficiency of transmission operations and reduces costs [5]. Inherent weather uncertainties in weather

could result in uncertainties in DTR and render system insecurity and thermal overloading. Probabilistic DTR modeling with percentile values or confidence levels is desired for reliable operation and analysis.

1.2 Major Contributions

This dissertation develops a set of novel methodologies to address challenges posed by the above particular problems.

1. A generic framework of day-ahead distribution-level load forecasting within the hierarchical structure is established to capture load characteristics of nodes at different levels, take advantage of the pattern similarities between a parent node and its child nodes, detect abnormalities, and provide high quality forecasts with low computational efforts. The new approach represents an effective way to forecast distribution-level load and would be helpful in the future smart grid.

2. A novel decomposition and coordination approach is developed for large day-ahead unit commitment with configuration-based combined cycle units. Our recently developed surrogate Lagrangian relaxation is significantly enhanced through adding quadratic penalties on constraint violations and a novel linearization to fully exploit exponential reduction of complexity with fast convergence. Enhancements on certain key aspects are also incorporated to fine tune the algorithm and improve the overall performance. The work is timely and critical to solve large UC problems, and can be extended to other complicated MILP problems in power systems and beyond.

3. A novel probabilistic dynamic thermal rating modeling is developed to consider both the topology of transmission lines and the inherent uncertainties in weather factors. The line thermal rating is obtained as the minimum of several span thermal capacities, which are forecast as random variables with means and standard deviations to capture the spatial nature and uncertainties. Expected values and important percentiles of the forecast ratings are approximately

extracted for system operations. The probabilistic approach would be beneficial for secure network operational management.

1.3 Organization of this Thesis

The rest of this thesis is organized as follows. Chapter 2 introduces an efficient load forecasting at the distribution level. Chapter 3 presents a novel decomposition and coordination approach for large day-ahead unit commitment with combined cycle units. Chapter 4 presents a probabilistic modeling of dynamic line ratings for overhead transmission lines.

References

[1] H. L. Willis, Spatial electric load forecasting, 2nd edition, New York: Marcel Dekker, 2005.

[2] Electric load forecasting. [Online] Available: http://www.quanta-technology.com/sites/default/files/doc-files/Load_Forecasting-12-01-13.pdf.

[3] Y. Chen, A. Casto, F. Wang, Q. Wang, X. Wang, and J. Wan, "Improving large scale day-ahead security constrained unit commitment performance," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4732-4743, 2016.

[4] Y. Chen and F. Wang, "MIP formulation improvement for large scale security constrained unit commitment," *Electric Power systems Research*, vol. 148, pp. 147-154, 2017

[5] E. Cloet and J. L. Lilien, "Uprating transmission lines through the use of an innovative real-time monitoring system," *IEEE PES 12th Int. Conf. on Transmission and Distribution Construction*, Operation and Live Line Maintenance (ESMO), May 2011.

Chapter 2

An Efficient Approach to Short-Term Load Forecasting at the Distribution Level

Short-term load forecasting at the distribution level predicts the load of substations, feeders, transformers, and possibly customers from half an hour to one week ahead. Effective forecasting is important for the planning and operation of distribution systems. The problem, however, is difficult in view of complicated load features, the large number of distribution-level nodes, and possible switching operations. In this Chapter, a new forecasting approach within the hierarchical structure is presented to solve these difficulties. Load of the root node at any user-defined subtree is first forecast by a wavelet neural network with appropriate inputs. Child nodes categorized as “regular” and “irregular” based on load pattern similarities are then forecast separately. Load of a regular child node is simply forecast as the proportion from the parent node load forecast while the load of an irregular child node is forecast by an individual neural network model. Switching operation detection and follow-up adjustments are also performed to capture abnormal changes and improve the forecasting accuracy. This new approach captures load characteristics of nodes at different levels, takes advantage of pattern similarities between a parent node and its child nodes, detects abnormalities, and provides high quality forecasts as demonstrated by two practical datasets.

2.1 Introduction

Power distribution systems deliver electricity from distribution substations to residential, commercial, and industrial consumers [1]. A distribution substation is fed from one or more transmission or sub-transmission lines and serves multiple feeders. Distribution transformers receive the power from one or more feeders and reduce the primary voltage to levels at which customers can use. In the distribution system when some branches are overloaded, there is a need to reconfigure the system by changing status of line switches to be open or closed [2]. These reconfigurations by switching operations can achieve load balance among distribution feeders, relieve overloading of the components and reduce system losses [3]-[5].

Short-term load forecasting at the distribution level predicts the load of substations, feeders, transformers, and possibly customers with a typical forecasting horizon ranging from half an hour to one week [1]. High quality load forecasting is important for the planning and operation of distribution systems. For instance, substation and feeder forecasts provide utilities with advanced warnings on potential substation and feeder overloading [5]. Customer load forecasting helps utilities schedule and dispatch community storage batteries to shave peak load in the smart grid environment [5], [6].

Forecasting the distribution-level load is much more difficult than forecasting a system-level load such as New England's load in view of the complicated load features, the large number of nodes, and the possible switching operations in distribution systems. Typically, load forecasts of a large area have high accuracy because the aggregated load is stable and regular, mainly resulting from the law of large numbers [7], [8]. However, the distribution-level load could be dominated by a few large customers such as industrial companies or schools [1], [5], and the load pattern may not be as regular as that of a large area. Moreover, considering the large number and the different load features at different distribution levels, usage of a unique forecasting model for all nodes may not be accurate. However, if an individual model were built for each node, it would be complicated and time-consuming for the system operation

and maintenance. In addition, due to the reconfigurations by switching operations, load may be temporarily switched from one feeder to another, which would severely change the distribution-level load profiles and affect the trend in a certain period. Without an advance notice to a load forecaster and a follow-up adjustment of forecasting methods, the forecasting power may be degraded.

To overcome the above difficulties, this Chapter presents a generic framework of day-ahead distribution-level load forecasting within the hierarchical structure. Generally, each node is fed from one line and the forecaster usually does not know switching operations in advance, therefore, the distribution-level load can be forecast within the hierarchical structure [9], [10] as depicted in Figure 2.1. Each node represents a distribution-level load and the load of a parent node is the aggregation of its child loads. In our forecasting framework, at any user-defined subtree, load of the root node is first forecast by an individual model. Dynamic node classification based on the load pattern similarities is then applied on each forecast day to categorize the child nodes as “regular” and “irregular”. Different load forecasting methods are developed regarding different types of child nodes. Load forecasts of lower-level nodes are obtained in the same manner by treating the current child node as a new parent node. The realized loads are examined through control ranges generated from forecast means and standard deviations to detect possible switching operations. Forecasting methods are then adjusted as needed to improve the forecasting accuracy for future days. The overall framework captures load characteristics of nodes at different levels, takes advantage of the pattern similarities between a parent node and its child nodes, detects abnormalities, and provides high quality forecasts with low computational efforts.

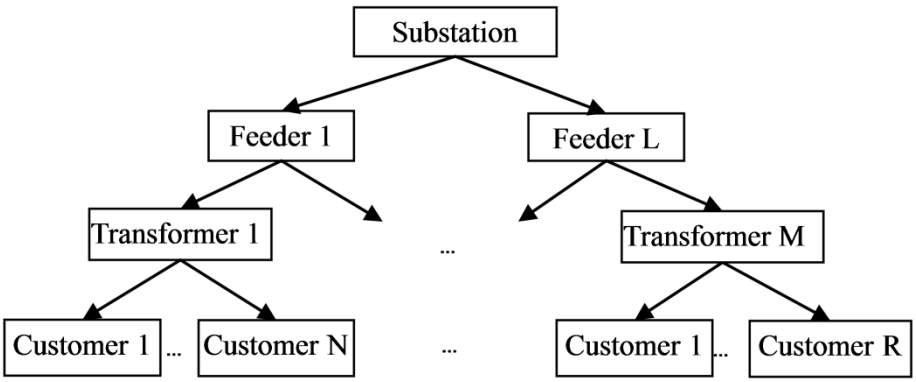


Figure 2.1. Hierarchical structure of the distribution-level load forecasting.

In Section 2.3, load forecasting for a root node, node classification, and load forecasting for different types of child nodes are presented. Load of a root node is first forecast by wavelet neural networks (WNNs) with selected inputs capturing load features and performing good predictions. With its child nodes classified as “regular” and “irregular” based on load pattern similarities, a different method is developed for each category. Load of a regular child node is directly forecast as a proportion of root load forecasts by load distribution factor (LDF). For an irregular node, correlation with a selected sibling node is taken into account and incorporated into an individual WNN model.

In Section 2.4, detecting switching operations and adjusting the forecasting methods are introduced to overcome the difficulties raised from feeder reconfigurations. Statistical Process Control (SPC) is used to monitor the actual load and detect abnormal changes according to load forecast means and standard deviations. If there is no switching operation, the actual load generally falls into a normal range. When a switching operation happens, actual load may exceed the normal range, causing significant changes to be caught by SPC rules. Because the abnormalities may affect the load trend and consequently degrade the forecasting accuracy, once a switching operation is identified, the forecasting methods will be adjusted as needed.

Two examples are provided in Section 2.5 to verify the effectiveness of our approach. Example 1 shows the load forecasting for one substation and six feeders, examining the effects of input selection, node classification, forecasting methods for regular and irregular nodes, and switching operation detection. Example 2 investigates the load forecasting of one substation with four feeders and smart meter-based customers. In both examples, our approach is compared with two naive benchmarks, two multiple regression models, and a simple neural network model. Numerical results show that our method outperforms all comparing models with high forecasting accuracy and low computational efforts.

2.2 Literature Review

2.2.1 Short-Term Load Forecasting for a Large Area

Different methods have been used for short-term load forecasting for a large area, including parametric and non-parametric regression models, Kalman filter, neural networks, and hybrid methods. Parametric regression models assume functional forms that describe relationships between load and affecting factors. The commonly used function models are explicit time functions, polynomial functions, autoregressive moving average (ARMA), Fourier series, and multiple linear regression (MLR). In contrast, non-parametric regression models do not take predetermined forms but are constructed according to information derived from the data. In Kalman filter, load is modeled in the state space formulation consisting of linear system state equations and measurement equations [11], [12]. The method is attractive because of the recursive property of Kalman filter and the standard deviations of forecasts obtained as byproducts. The main difficulties in Kalman filter are the state selection and model identification.

From the late 1980's, much research has been studied on applying artificial intelligence techniques to load forecasting. Among these, neural networks (NNs) have been widely used because of their strong ability to approximate the nonlinear function through learning historical data [13]. The NNs have also

been combined with other methods to improve the prediction power. A combination of radial basis function neural networks and adaptive neural fuzzy inference was established in [14] to forecast load in real-time price environments. A similar day-based back propagation neural network was developed in [15] to forecast the next day load. In this method, similar day load is selected as NN inputs based on the similarity between the forecast day's predicted weather and the historical days' weather. The NNs are commonly trained by back propagation algorithm. As a first-order steepest decent method, back propagation suffers from slow convergence and may not be efficient for nonstationary process. The extended Kalman filter (EKF) has been used to train a NN by treating weights of the network as the state of a nonlinear dynamic system [16] because of its strong tracking capability. Since it is a second-order algorithm, fast convergence is expected. Nevertheless, using EKF in load forecasting may require much computational effort considering the high dimensionality of the weights involved. Decoupled EKF, which is a simplified form of EKF, reduces the computational time by ignoring some dependency of weights such that the weight covariance matrix is block diagonal. Among possible decoupling strategies [17], node-decoupled EKF, in which each weight group is composed of a single node's weights, is straightforward and applied to simplify EKF. A NN-based market clearing price forecaster with node-decoupled EKF presented in [18] showed good prediction performance and a significant decrease of the computational time. The above methods shed insights on model selection and affecting factor identification for load forecasting at the distribution level.

2.2.2 Short-Term Load Forecasting at the Distribution Level

Many methods have been reported on short-term load forecasting at the distribution level. Some researchers focus on forecasting one particular substation or feeder, and others forecast a large number of distribution-level loads together.

Load forecasting of a substation or a feeder encounters high errors as a result of the complicated load features [19]. Different load patterns of small regions within a large geographic area were presented in

[20]. Load diversity was quantified in [21] to represent levels at which regional load affected the overall system load. In [22], a hybrid method composed of a forecast-aided state estimator and a NN was presented for substation load forecasting. To better track the nonstationary substation load, outputs of the state estimator were used as initial forecasts and fed to NN to generate final forecasts. This hybrid method saved computational effort compared with pure NN methods, however, considering the large number of distribution-level nodes, using an individual model for each node may not be effective for system operations and maintenance.

The following presents general models to forecast loads of a large number of nodes. In [23], load features of 245 substations from a national grid were analyzed. Correlation coefficients with weather factors and day indices showed different types of sensitivities across substations. Periodic autoregressive models were used for short-term substation load forecasting with monthly, weekly and the intra-daily patterns modeled. In [24], a semi-parametric load forecasting model was developed for over 2000 substations in French grid. Using a unique regression-based model to forecast all time series saves the computational efforts. However, it is difficult to capture the characteristics of loads at different levels. Furthermore, the above methods treated each small area or grid component as a separate entity, and forecasts were produced without any regards to any information available from areas outside.

Hierarchical load forecasting is an approach in which load forecasts at different hierarchy levels are connected. A hierarchical forecasting model developed in [9] provided load forecasts for system, areas, zones, and substations. The NN-based forecasting engines were associated at any user-defined nodes. Load forecasts for other nodes were obtained using aggregation and load distribution factor (LDF). Conceptually LDF is the ratio of a child load to its parent load and can be calculated in several ways. For instance, in [25], LDF for each substation was forecast individually through a general regression neural network model. In [10], two types of LDFs were introduced: a short LDF was calculated by the latest data and used for the next time instance while a long LDF was calculated based on the latest daily data and

used for the next day. The above work simplified forecasting procedures for each node with a light computational effort. Considering that the load pattern of a child node could be significantly different from that of its parent node, it is not proper to forecast all child nodes by LDF. Moreover, pattern similarities may change over time. Thus, a dynamic node classification method is required to categorize the child nodes and capture changes of the load patterns.

Effects of switching operations on distribution-level load forecasting have been reported in [26] and [27]. A two-stage bad data identification method was developed in [27] to retrieve the historical trend of load and to improve the bus load forecasting by identifying and restoring inaccurate measurements and abnormal disturbance. A synergistic integration of Statistical Process Control (SPC) and Kalman filter was presented in [28] to detect faults of chillers and cooling coils by monitoring system parameters. The SPC was applied to evaluate variations of parameter predictions while Kalman filter was used to provide predictions and adaptive SPC control limits. The above detection methods captured abnormal changes effectively and could be adopted to detect switching operations.

2.3 Neural Networks and Load Distribution Factors

Our new forecasting approach is presented in this section. Subsection 2.3.1 presents a WNN-based load forecasting method for a root node. Subsection 2.3.2 describes a criterion to classify child nodes as “regular” and “irregular” based on load pattern similarities. The LDF method to forecast load of a regular node is expressed in Subsection 2.3.3. Subsection 2.3.4 introduces the WNN with appropriate inputs for irregular nodes.

As illustrated in Figure 2.2, load of the root node Feeder 3 in a given subtree is first forecast by the WNN. According to the results of node classification, the forecasts are distributed to regular nodes (transformers 1 and 3) by using LDF while the load of the irregular node transformer 2 is forecast by

using WNN. Load forecasts of customers 1-3 are then obtained in the same manner by treating transformer 2 as a new parent node.

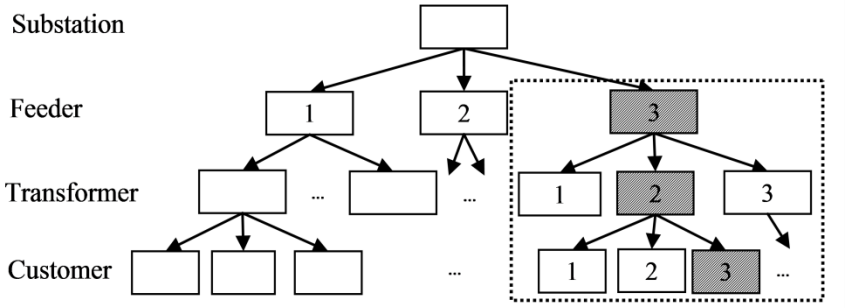


Figure 2.2. An illustration of the novel distribution-level load forecasting method. In a subtree where Feeder 3 is the root node, nodes with shadowing are forecast by WNNs and the rest are forecast by LDFs.

2.3.1 Load Forecasting for a Root Node

Previously, we have developed a WNN-based load forecasting method for large areas [15], [29]. The structure consists of similar day-based input selection, multi-level wavelet decomposition, and individual neural networks for different frequency components. The forecasting model for a root load follows this structure with modifications on input selection and the learning algorithm as depicted in Figure 2.3. Inputs to WNNs include similar day load, previous day load, forecast weather, and day of week index while outputs are forecast loads of the next day at all time instances. The wavelet technique is used to decompose the data into three orthogonal components at different frequencies: Low-low (LL), Low-high (LH) and High (H). This process helps capture load features of individual components. Results of individual NNs are then summed to form the final forecasts.

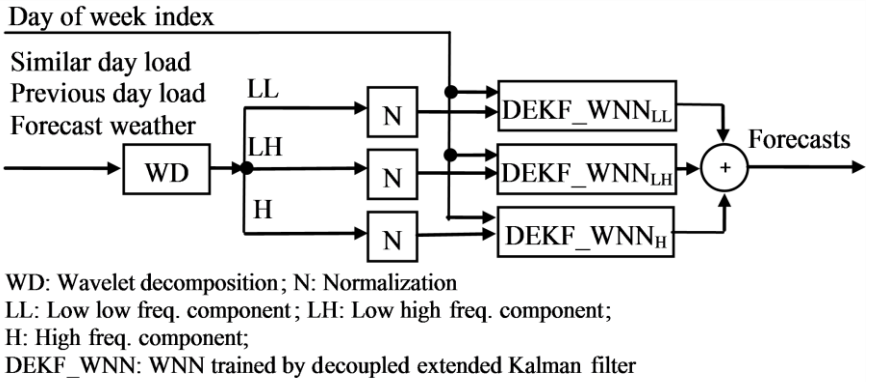


Figure 2.3. The WNN-based load forecaster for a root node.

2.3.1.1 Input Selection

To effectively capture features of the distribution-level load and forecast the next day load, inputs are properly selected. Day of week index is an important input factor because different days of the week have different load curves. Beyond that, weather is the major driver for load. Following [15], predicted wind-chill temperature and humidex on the next day are selected as weather input variables in WNNs. As shown in Figure 2.4, approximate piecewise linear relationships exist between substation load and selected weather factors. Other weather factors such as wind speed, which has a highly nonlinear pattern and a weak correlation with the load, are not selected.

In the method we previously developed for system-level load forecasting [15], similar day load is selected based on weather similarity and day of week index. Distribution-level load, which could be dominated by a few large customers, may vary with similar weather conditions and the same day of week index. Therefore, only considering weather similarity may not be sufficient. Typically, if the day before a weather-similar day also has a similar load curve with the day before a forecast day, the selected similar day load would better represent the forecast day load. The criteria of similar day selection (4) in [15] are thus modified according to

$$\min_{i} \sum_{t=1}^{T} \left| W^{f}(t) - W^{i}(t) \right| + \alpha \sum_{t=1}^{T} \frac{\left| L^{f-1}(t) - L^{i-1}(t) \right|}{\bar{L}}, \; i \in \Theta, \tag{2.1}$$

where subscripts f and i, respectively, denote a forecast day and a historical day in the historical set Θ; subscripts f-1 and i-1, respectively, denote the day before the forecast day f and the day before the historical day i; W represents the weather factor under consideration, i.e., wind-chill temperature if tomorrow is a winter day, and humidex if tomorrow is a summer day; L represents the load and T is the number of time instances during one day. $\bar{L}$, which is the average value of the historical load, is used to scale the magnitude of load differences. The weight α of the load difference term is determined by the following minimization process on all historical days:

$$\min_{\alpha} \frac{1}{D} \sum_{d=1}^{D} \sum_{t=1}^{T} \left| SD^{d}(\alpha, t) - L^{d}(t) \right|, \; d \in \Theta, \; \alpha \in N_{+}, \tag{2.2}$$

where $SD^{d}(\alpha, t)$ denotes the selected similar day load with parameter α of a historical day d at time instance t; $L^{d}(t)$ denotes the actual load of day d at time instance t; and D is the number of historical days.

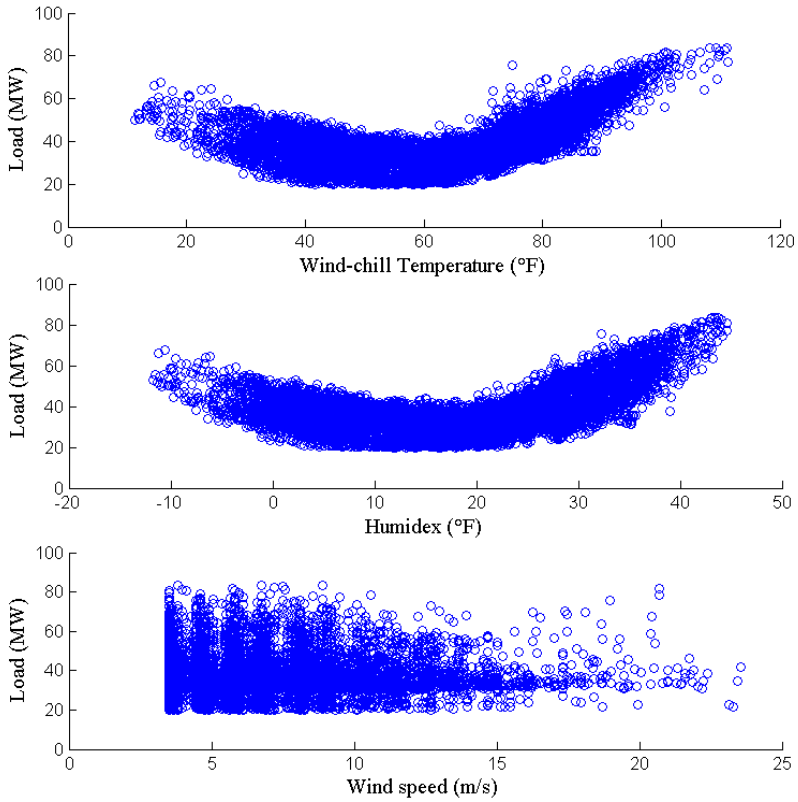


Figure 2.4. Load versus wind-chill temperature, humidex, and wind speed.

Profiles of a substation actual load, original and modified similar day loads, and previous day (Day-1) load are depicted in Figure 2.5. The mean absolute percent error (MAPE) and correlation coefficient ρ with respect to the actual load are used to evaluate these load input variables according to

$$MAPE(L_A, L_V) = \frac{1}{n}\sum_{i=1}^{n}\frac{|L_A(i) - L_V(i)|}{L_A(i)} \times 100\% \tag{2.3}$$

and

$$\rho_{L_A,L_V} = \frac{\sum_{i=1}^{n}\left(L_A(i) - \bar{L}_A\right)\left(L_V(i) - \bar{L}_V\right)}{\sqrt{\sum_{i=1}^{n}\left(L_A(i) - \bar{L}_A\right)^2}\sqrt{\sum_{i=1}^{n}\left(L_V(i) - \bar{L}_V\right)^2}}, \tag{2.4}$$

where L_A denotes the actual load, L_V represents the load input variable to be examined, n is the number of historical samples, $\bar{L}_A$ and $\bar{L}_V$ are the sample means. The MAPE measures the closeness between the input load and the target load while the correlation coefficient measures the association between the input load and the target load. As summarized in Table 2.1, modified similar day load has the lowest MAPE and the highest correlation coefficient compared with other two load input variables. This indicates that the modified similar day load could better represent the load of a forecast day compared with the other two load input variables. Meanwhile, to anchor the selected similar day load and provide an initial status of the next day load, previous day load is supplemented to the modified similar day load. Numerical testing of different combinations of weather and load input variables are provided in Section 2.5 to verify the above selection.

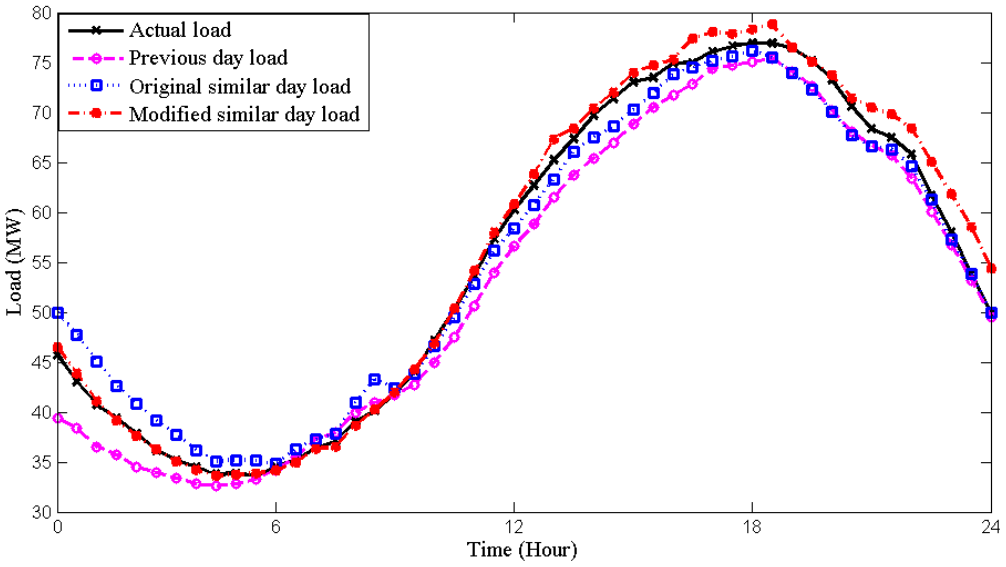


Figure 2.5. Load curves of actual load, previous day load, original similar day, and modified similar day load.

Table 2.1 Comparisons of Load Input Variables with Target Output Load

Method	**Previous day load**	**Original similar day load**	**Modified similar day load**
MAPE (%)	9.07	7.67	6.21
Correlation coefficient	0.89	0.92	0.94

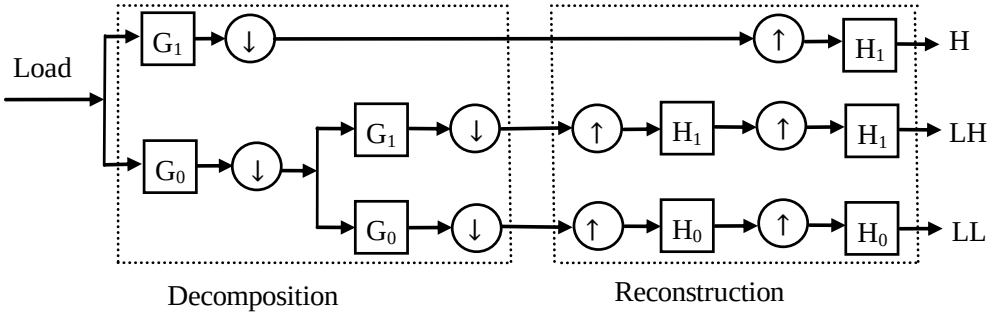


Figure 2.6. Multiple-level wavelet decomposition scheme (Load = LL + LH +H)

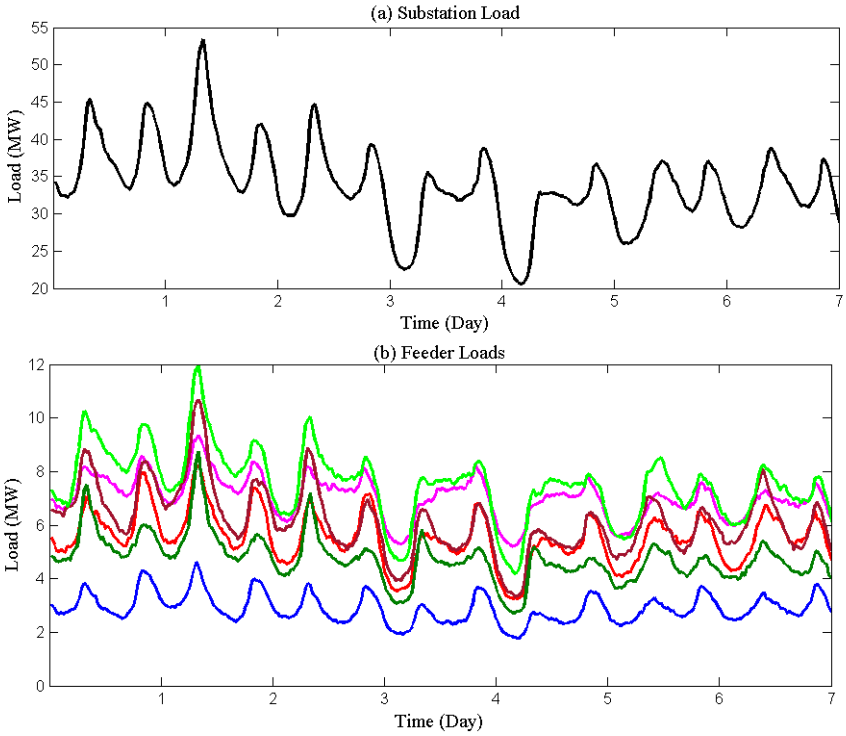


Figure 2.7. Load profiles of (a) a substation and (b) six feeders during a week from March 5, 2012 (Monday) to March 11, 2012 (Sunday).

2.3.1.2 Wavelet Neural Network

For Combination of wavelet transform and NNs has been successfully used in load forecasting [15], [30], [31]. Unlike Fourier transform, which represents the signal as a sum of sinusoids localized in frequency only, wavelet transform uses basis functions which contain both time and frequency information. It is thus appropriate to use wavelet transform to deal with signals with nonstationary characteristic through multi-resolution analysis.

Wavelet transform can be implemented by a filter bank presenting decomposition and reconstruction stages as shown in Figure 2.6. In the decomposition stage, approximation and detail coefficients of an input signal are produced by convolving with filters (G_0, G_1) and then by down-sampling. An 'approximation' holds the general trend of the original signal, whereas a 'detail' depicts high frequency component of it. In the reconstruction stage, wavelet coefficients are padded zeros (up-sampling) to

recover the data length and are then convolved with reconstruction filters (H_0, H_1). Filters G_0, G_1, H_0, and H_1 have to satisfy perfect reconstruction and orthogonality [32]. The input load is thus broken into low and high frequency components. A multilevel decomposition process can be achieved by successively decomposing the approximations. In this Chapter, a two-level wavelet composition is adopted. Thus, load and weather data are decomposed into LL, LH and H frequency components as the scheme presented. When these decomposed components are fed into individual NNs trained by node-decoupled EKF, the same forecasting quality is assumed for each component. The three forecasts are therefore summed with equal weights to obtain the final forecast.

Daubechies (Db) wavelets are selected in our method because they are a family of orthogonal wavelets, and will not cause information loss in the frequency domain. The Db members tested are Db2-Db20 (even index only), in which the index number refers to the filter length. To choose a good decomposition level and filter window length, extensive experiments are conducted [29]. Two-level decomposition with Db4 is found to be the best among levels from zero to three and DB index members from 2-20. Determination of the above parameters as well as the parameters in NNs such as the number of hidden neurons is through training, validation, and test processes in a three-way data split [33]:

Step 1: Divide the data into training, validation and test sets;

Step 2: Select one parameter (e.g., number of hidden neurons) and set values of other parameters to nominal levels.

Step 2. 1: Initialize the selected parameter;

Step 2.2: Train WNN using the training set;

Step 2.3: Evaluate the model on the validation set by calculating the validation set error;

Step 2.4: Tune the parameter and repeat steps 2.2 and 2.3;

Step 2.5: Select the parameter value which results in the minimum validation set error;

Step 3: Repeat step 2 to determine other parameters;

Step 4: Select the best model and train it using data from the training and validation sets;

Step 5: Assess the final model using the test set.

2.3.2 Node Classification

Load pattern of a child node generally follows the pattern of its parent node. As weekly load profiles of a substation and its six feeders shown in Figure 2.7, most feeders have similar load patterns with that of the substation. Nevertheless, there also exist deviations on particular days (e.g., Day 3 and Day 4) when load patterns of some feeders are different from that of the substation. The load pattern similarities motivate forecasting child nodes by the proportion from the parent load forecast. However, for those child nodes whose loads significantly vary from the parent load, proportion is not suitable and individual models are needed. To identify which category that a child node belongs to and select a proper forecasting method, a dynamic node classification method is developed.

Similarity between two time series is commonly identified by distance matrix [34]-[36]. In [34], similarity matching of two observed series was based on the distance of wavelet coefficients after decomposing the data series. In [35] and [36], simple Euclidean distance was used to investigate the pattern similarities. In our problem, considering the different magnitudes of child nodes, Euclidean distance between the normalized child load and parent load is calculated as a pattern similarity index. Furthermore, since loads having the same day of week index are likely to have similar patterns, the distance between a child node *i* and its parent node on day *k* is estimated as the averaged distance of past *S* weeks having the same day index with the forecast day as expressed in

$$d(i,k)=\frac{1}{S}\sum_{s=1}^{S}\sqrt{\sum_{t=1}^{T}\left(L_{Nor,C}(i,k-7s,t)-L_{Nor,P}(k-7s,t)\right)^2} \qquad i=1,...,N, \quad t=1,...,T, \qquad (2.5)$$

and

$$L_{Nor}(k,t)=\frac{L(k,t)-\min\limits_{t=1,\ldots,T}(L(k,t))}{\max\limits_{t=1,\ldots,T}(L(k,t))-\min\limits_{t=1,\ldots,T}(L(k,t))}, \tag{2.6}$$

where N is the number of child nodes, $L_{Nor,C}(i,k,t)$ and $L_{Nor,P}(k,t)$ are the normalized loads of child node i and the parent node at time instance t on day k, respectively.

Distances between all child nodes and the parent node are calculated using (2.5) and (2.6). Node classification is then determined for the forecast day k according to:

$$node\,i=\begin{cases} regular\,node & if\ d(i,k)\le d_s(k) \\ irregular\,node & if\ d(i,k)>d_s(k) \end{cases}, \tag{2.7}$$

where d_s is a distance threshold. On each forecast day, a child node is defined as "regular" if d is smaller than the threshold; otherwise, it is defined as "irregular". The threshold d_s is initialized as the average of historical distances, and then tuned through training and validation data sets.

This distance-based classification method can be treated as a simplified form of K-means clustering in which the normalized parent load is the cluster center. Compared with the standard K-mean clustering with two clusters, our classification calculates the distance between child and parent nodes and indicates how much the child loads follow the parent load. Numerical results for the estimated distances, actual distances, and the threshold distance are shown in Section 2.5 to verify and evaluate the classification.

2.3.3 Load Forecasting Method for Regular Nodes

To forecast the load of a regular node i on day k, LDFs are estimated by the averaged LDFs of the past S weeks having the same day of week index with the forecast day as in

$$LDF(i,k,t)=\frac{1}{S}\sum_{s=1}^{S}\frac{L_C(i,k-7s,t)}{L_P(k-7s,t)}. \tag{2.8}$$

The LDFs are updated on each forecast day and this dynamic feature helps capture changes of the proportions. Load forecasts of a regular child node are then obtained simply as

$$\hat{L}_C(i,k,t) = \hat{L}_P(k,t) \cdot LDF(i,k,t), \tag{2.9}$$

where $\hat{L}_C(i,k,t)$ and $\hat{L}_P(k,t)$ are load forecasts of child node *i* and its parent node, respectively.

2.3.4 Load Forecasting Method for Irregular Nodes

For irregular nodes, since their load patterns could be significantly different from that of the parent node, LDF is not suitable. To better capture the complicated load features and estimate the nonlinear relationships between affecting factors and the target load, WNN-based forecasting model is used.

In spatial load forecasting, correlations with neighboring regions and information available outside the target area are used to improve the forecasting accuracy [5]. Using information available from a neighboring area to predict a target area has achieved improvements in load forecasting [37], wind power forecasting [38] and solar power forecasting [39]. As discussed above, in addition to the inputs considered for a root node, load from correlated sibling nodes would help WNNs capture the load features of irregular nodes. Since the number of correlated nodes could be large, and if all are considered, algorithm complexity will increase and forecasting accuracy may even degrade. Our idea is to select one key sibling node and use the previous day load of this selected node as additional inputs to WNNs.

2.4 Detection Switching Operations

This section deals with detecting switching operations and adjusting forecasting methods after an identified switching operation. Subsection 2.4.1 presents an SPC-based method to detect switching operations. In Subsection 2.4.2, the methods for adjusting LDFs and WNNs are described.

2.4.1 An SPC-based Method for Detecting Switching Operations

Statistical Process Control (SPC) is used to monitor the actual load and detect abnormal changes according to load forecast means and standard deviations. If there is no switching operation, the actual load generally falls into a normal range. When a switching operation happens, actual load may exceed the normal range, causing significant changes to be caught by SPC rules. This switching operation detection method adopts the SPC-based detection idea and the SPC control rule from [28], in which a synergistic integration of SPC and Kalman filter was presented to detect the faults of chillers and cooling towers.

Load forecast errors are commonly assumed to be Gaussian distributed [40]. Under this assumption, actual load falls into the one-sigma range with approximate 68% probability and two-sigma range with 95% probability without switching operations. In [28], two-sigma range is used as the adaptive SPC control limit to guarantee 95% Gaussian coverage. Similarly, a switching operation in the distribution system is detected if *n* back-to-back points fall outside of the two-sigma range and the points are either all above upper limits or all below lower limits. This is because a switching operation typically keeps the trend and lasts for a certain period. The number of points, *n*, is set to be 3 so that a potential switching operation is detected at the 99.99% confidence level according to (10) in [28].

For nodes forecast by WNN, standard deviations of the forecasts can be directly obtained from the diagonal elements of the innovation covariance matrix in decoupled EKF [18]. For regular nodes forecast by LDFs, standard deviations are derived approximately from those of their parent nodes. Assume that the load forecasting error of a parent node is Gaussian with standard derivation $\sigma_P(k, t)$ on day k at time instance t. The load forecasting error of its regular child node i is then approximately obtained as a Gaussian with standard derivation proportioned by LDF from $\sigma_P(k, t)$

$$\sigma(i,k,t) = LDF(i,k,t)\sigma_P(k,t) . \qquad (2.10)$$

Distributions of forecasting errors from WNN and LDF are analyzed to verify the normal assumption by Quantile-Quantile plots of the errors and sigma coverage rates of the actual load as demonstrated in the numerical testing.

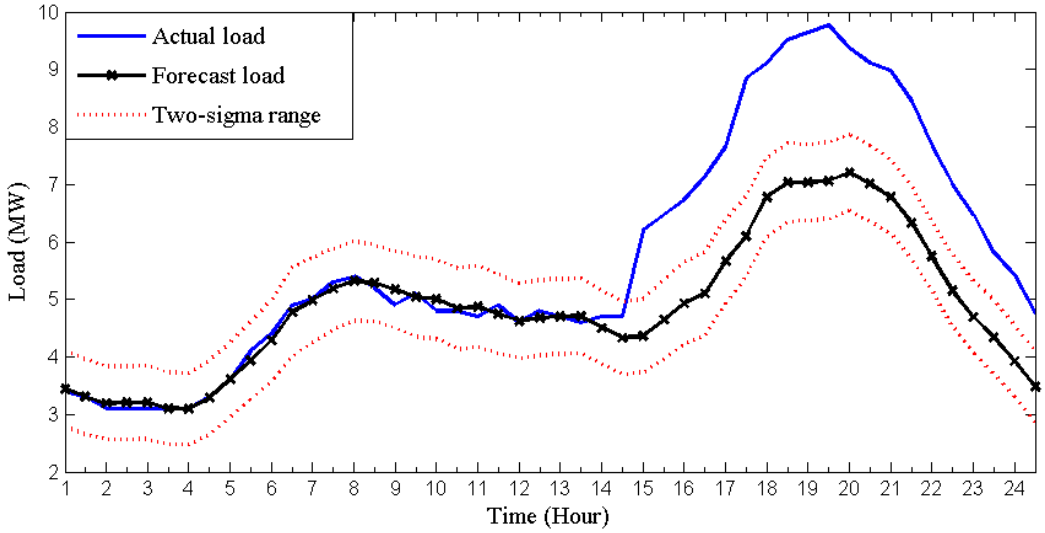


Figure 2.8. Dynamic two-sigma range as SPC control limits for detecting switching operations.

2.4.2 Adjusting LDFs and WNNs

After the actual load of a forecast day is available, the load is examined using the SPC control rule as discussed above. If no switching operation is detected and no earlier switching operation is identified, regular forecasting processes will be followed for the next day. If a switching operation is detected and the load is switched back to the normal range, i.e., the abnormality does not last long, regular forecasting processes will be followed.

If a switching operation is detected and it lasts until the end of the current day as shown in Figure 2.8, adjusting LDFs or WNNs is needed. For a regular node, after such a switching operation, we keep on using LDF but the calculation in (8) is based on the latest new data only without the requirement of using data with the same day index. When the length of the new data collected is more than *S* weeks, original LDF is used. However, after a switching operation is identified, a regular node may not maintain the

regular pattern. Thus, WNN is started at the background with the new data until it is shown that the new pattern is regular, or WNN is ready to forecast.

For an irregular node, after switching operations, we keep on using WNN with weights updated by the new data. Since the irregular node may become regular, LDFs are started with the new data at the background until it is shown that the new pattern is irregular, or when LDFs begin to produce good predictions.

2.5 Numerical Results

The above method has been implemented in MATLAB on an Intel Core 2.20GHz personal computer. The forecasting performance is evaluated by using the standard mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|\hat{L}(t) - L_A(t)\right|}{L_A(t)} \times 100\% . \tag{2.11}$$

If the denominator $L_A(t)$ in (11) is close to zero, mean absolute error (MAE) is used:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{L}(t) - L_A(t)\right| . \tag{2.12}$$

Two practical datasets are tested in Examples 1 and 2, respectively. For both examples, one-year load data in 30-minute interval and hourly weather data are collected. The test set is the last month data, the validation set is one month prior to the test month, and the rest data are used for WNN training. When the validation error (MAPE/MAE) increases for five iterations, the training is stopped, and the weights resulting in the minimum validation error are stored. Parameters need to be set in WNNs, node classification, and LDF calculation are determined based on training and validation processes as described in Subsection 2.3.1.

Since only actual weather data are available, in the training period, actual weather data are used as WNN inputs whereas in the validation and testing periods, actual weather data plus a Gaussian noise $N(0, 3)$ are used as weather predictions. The forecasting performance of using this predicted weather is expected to be close to that in the real applications.

2.5.1 Example 1

This example demonstrates load forecasting for a substation and six feeders located in a large city in North Carolina, United States. Load and weather data collected are from August 01, 2011 to July 24, 2012. The training period is from August 2011 to May 2012, the validation period is June 2012, and the test period is July 01-24, 2012.

Seven cases are presented below. Case 1 demonstrates the learning and generalization capability of WNN using one-year data for substation load forecasting. Case 2 shows the values of input selection of WNN for substation load forecasting. Case 3 examines the dynamic feeder classification. Case 4 shows load forecasting for regular feeders. Case 5 compares WNN and LDF for irregular feeders. Case 6 evaluates the normal assumption for load forecasting errors and demonstrates the significance of switching operation detection. In Case 7, the performance of our method is compared with two naïve benchmarks, two multiple linear regression models, and one NN model.

Case 1: The training and validation processes of each frequency component for substation load are described in this case. The inputs selected for WNN include similar day load, previous day load, forecast day weather predictions (wind-chill temperature and humidex), and day of week index. The training and validation errors are evaluated by MAPE for LL frequency component and MAE for LH and H components. As shown in Figure 2.9, for LL frequency component, both learning and validation processes converge quickly after several iterations. For LH and H frequency components, because the components are volatile, convergences are not as smooth as for the low frequency component. The validation set stops the training at a specific number of iterations if further five iterations on training data

degrade the network generalization ability on the validation set. The numbers of iterations to meet the minimum validation errors are 22, 10, and 7 for LL, LH and H frequency NNs, respectively.

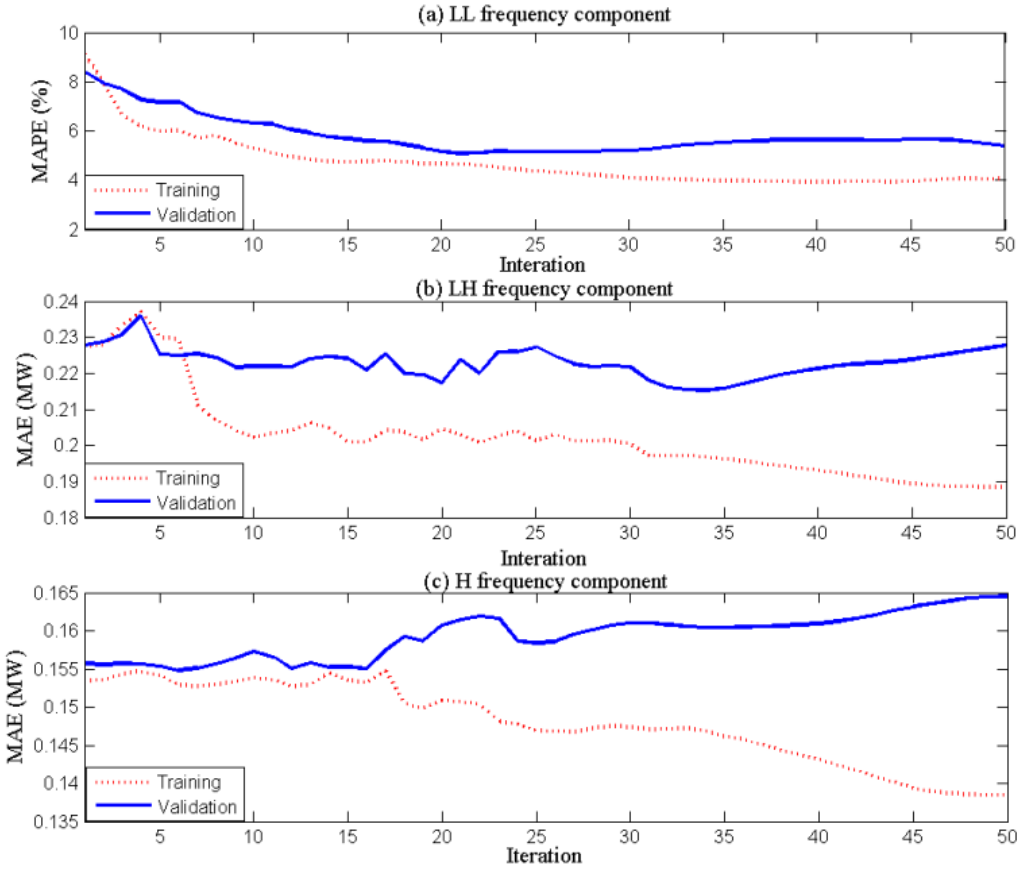


Figure 2.9. Training and validation errors with the number of training iterations for each frequency component.

Case 2: This case demonstrates the benefits of input weather and load variable selection for WNN to forecast the substation load. In our method, the inputs selected for WNN to forecast the load of a root node include similar day load, previous day load, forecast day weather predictions (wind-chill temperature and humidex), and day of week index.

1). *Weather input variables*: WNNs using different combinations of weather input variables are compared. Other inputs include similar day load, previous day load, and day of week index. MAPEs presented in Table 2.2 show the values of selected wind-chill temperature and humidex.

2). *Load input variables*: WNNs using different combinations of load input variables are compared. Other inputs include weather predictions (wind-chill temperature and humidex), and day of week index. The MAPEs presented in Table 2.3 show the values of selected load inputs.

3). *Actual weather verse predicted weather*: WNN using the actual weather and predicted weather on forecast days are compared. Results in Table 2.4 show that with adding a noise on the actual weather, the forecasting MAPE increases compared with using actual weather. This is reasonable because using actual weather without uncertainties could result in lower MAPE. However, the performance of using weather data with uncertainty is expected to be close to the performance in the real applications.

TABLE 2.2 Number of Hidden Neurons (N_H) and MAPE (%) for Substation July, 2012 Load

	M1	**M2**	**M3**	**M4**	**M5**
N_H LL/LH/H	20/15/15	20/15/15	20/15/15	23/18/18	25/18/18
MAPE	5.46	5.45	5.91	5.17	5.93

M1: With wind-chill temperature; M2: With humidex; M3: With wind speed;
M4: With both wind-chill temperature and humidex;
M5: With wind-chill temperature, humidex and wind speed.

TABLE 2.3 MAPE (%) for Substation July, 2012 Load (Case 2-2 in Example 1)

	M1	**M2**	**M3**	**M4**	**M5**
N_H LL/LH/H	20/15/15	20/15/15	20/15/15	23/18/18	23/18/18
MAPE	5.88	5.87	5.62	5.44	5.17

M1: With previous day load; M2: With original similar day load;
M3: With modified similar day load;
M4: With both previous day load and original similar day load;
M5: With both previous day load and modified similar day load.

TABLE 2.4 MAPE (%) for Substation July, 2012 Load (Case 2-3 in Example 1)

With Actual Weather	With Forecast Weather
4.98	5.17

Case 3: The node classification process is presented in this case. In Table 2.5, the first row shows the actual averaged distances calculated during the validation period June 2012 while the rest are the estimated distances on the last two weeks of the test period. The distance threshold is set as 0.50, which is determined based on the historical actual distances and training-validation process. It can be seen that Feeders 1, 2, 3 and 5 are identified as "regular" nodes for all test days and they are forecast by LDF. Feeder 4 is identified as an "irregular" node for most of the test days. Even though the estimated distances on some days (e.g., July 16 and 17 with $d < 0.5$) show regularity of Feeder 4, since this regularity does not keep long, WNN is still used until a long-term regularity is identified. As for Feeder 6, it has only one day identified as an irregular node during all test days. In view of the fact that the averaged distance during validation period shows the regularity of this node, and the irregularity did not keep long, Feeder 6 is therefore treated as a regular node and forecast by LDF for all test days. Thus five of six feeders are forecast by LDFs, which indicates a low computational effort. Actual distances during the last two weeks of the test period are also calculated for verification as shown in Table 2.6. The actual distances are generally consistent with the estimated and this demonstrates the overall irregularity of Feeder 4 and the regularity of other feeders.

TABLE 2.5

Estimated Distances of the Normalized Loads between Six Feeders and Substation (Case 3 in Example 1)

Data	F1	F2	F3	F4	F5	F6
June 2012 (Actual)	**0.29**	**0.26**	**0.28**	**0.63**	**0.44**	**0.39**
7-11-2012	0.31	0.30	0.24	**0.51**	0.33	0.41

7-12-2012	0.28	0.37	0.30	**0.58**	0.38	0.47
7-13-2012	0.26	0.31	0.31	**0.63**	0.48	0.50
7-14-2012	0.31	0.23	0.31	**0.69**	0.48	0.48
7-15-2012	0.22	0.24	0.27	**0.65**	0.45	0.48
7-16-2012	0.21	0.22	0.20	0.44	0.49	0.36
7-17-2012	0.20	0.22	0.19	0.30	0.26	0.26
7-18-2012	0.30	0.27	0.27	**0.58**	0.34	**0.51**
7-19-2012	0.23	0.26	0.32	**0.55**	0.40	0.49
7-20-2012	0.22	0.19	0.28	**0.54**	0.39	0.46
7-21-2012	0.24	0.21	0.27	**0.56**	0.40	0.45
7-22-2012	0.23	0.20	0.23	**0.54**	0.37	0.42
7-23-2012	0.22	0.21	0.19	0.45	0.37	0.37
7-24-2012	0.21	0.22	0.20	0.32	0.26	0.28

TABLE 2.6

Actual Distances of the Normalized loads between Six Feeders and Substation (Case 3 in Example 1)

Data	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**
June 2012	**0.29**	**0.26**	**0.28**	**0.63**	**0.44**	**0.39**
7-11-2012	0.30	0.49	0.35	**0.70**	0.42	**0.60**
7-12-2012	0.33	0.26	0.38	**0.84**	0.49	**0.55**
7-13-2012	0.25	0.29	0.35	**0.99**	**0.63**	**0.60**
7-14-2012	0.23	0.23	0.28	**0.68**	0.44	0.43
7-15-2012	0.17	0.23	0.14	0.38	0.46	0.31
7-16-2012	0.18	0.24	0.23	0.26	0.19	0.23
7-17-2012	0.21	0.22	0.32	0.50	0.40	0.47
7-18-2012	0.17	0.13	0.29	0.45	0.33	0.47
7-19-2012	0.28	0.21	0.27	**0.52**	0.37	0.47
7-20-2012	0.21	0.20	0.23	**0.51**	0.38	0.41
7-21-2012	0.24	0.23	0.22	**0.59**	0.33	0.41
7-22-2012	0.21	0.24	0.20	**0.53**	0.49	0.38
7-23-2012	0.24	0.21	0.20	0.20	0.21	0.23
7-24-2012	0.32	0.28	0.27	**0.53**	0.44	**0.53**

Case 4: Load Forecasting for regular nodes Feeders 1, 2, 3, 5 and 6 by using LDF and individual WNN are compared. The results summarized in Table 2.7 show that the LDF captures load features and provides competitive load forecasts for regular nodes compared with the results from WNNs.

TABLE 2.7

MAPE (%) for Regular Feeders July, 2012 Load (Case 4 in Example 1)

	LDF	**WNN**
Feeder 1	5.63	5.49
Feeder 2	5.73	5.95
Feeder 3	5.91	5.77
Feeder 5	5.35	6.05
Feeder 6	5.44	5.22

Case 5: Load forecasting performance for an irregular node is shown in this case. Irregular node Feeder 4 is forecast by WNN with considering correlations with the selected sibling node Feeder 3. Three other approaches are compared: using WNN without considering correlations with other nodes, using WNN with the substation load as additional inputs, and using LDF method. Forecasting MAPEs are summarized in Table 2.8. For Feeder 4, our method (M1) produces the lowest MAPE 7.25 as compared with other two WNNs and LDF.

TABLE 2.8

MAPE (%) for Feeder 4 July, 2012 Load (Case 5 in Example 1)

M1	**M2**	**M3**	**M4**
7.25	7.38	7.45	7.72

M1: Using WNN with a correlated sibling load as additional inputs;
M2: Using WNN without any load inputs outside this node;
M3: Using WNN with the substation load as additional inputs;
M4: Using LDF.

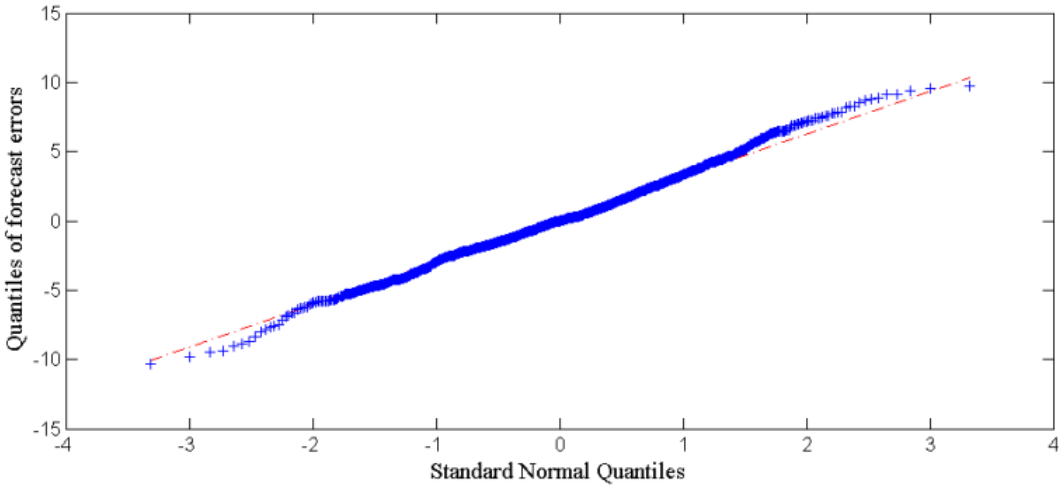


Figure2.10. Quantile-Quantile plot of substation load forecasting errors versus the standard normal.

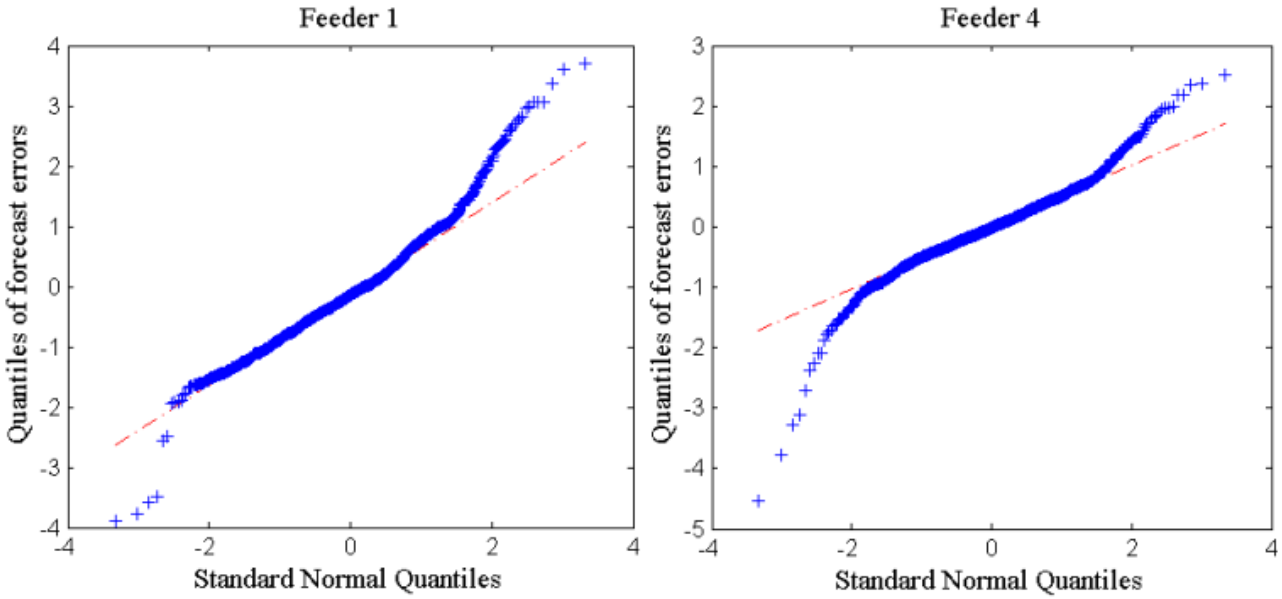


Figure 2.11. Quantile-Quantile plots of forecasting errors versus the standard normal for a regular load Feeder 1 and an irregular load Feeder 4.

Case 6: This case verifies the normal assumption of the forecasting errors and demonstrates the switching operation detection. As shown in Figure 2.10, Quantile-Quantile plot of the substation load forecasting errors clearly shows heavier tails than the Gaussian. The Quantile-Quantile plots of load

forecast errors for a regular node Feeder 1 and an irregular load Feeder 4 are also depicted in Figure 2.11. Both have heavier tails than Gaussian. However, if the top and bottom tails are removed, the remaining errors follow a normal distribution. The standard deviation (STD) of the prediction, one-sigma and two-sigma actual load coverage rates calculated for substation and feeders are shown in Table 2.9. The one-sigma coverage rates range from 69.18% to 81.07%, which are slightly larger than 68% under the Gaussian assumption. The two-sigma coverage rates of all feeders are lower than the value 95% under Gaussian distribution. This is mainly because of the abnormal consumption occurred on July 2 for Feeder 1 and on July 8 for almost all feeders. These changes are caught by the two-sigma rule as shown in Figure 2.12 for Feeder 1. The two-sigma coverage rates of the actual load are thus lowered. Because the abnormal load was switched back to the normal ranges, no adjustments are made in this case.

TABLE 2.9
MAPEs (%), MAEs (MW), STDs (MW), One-sigma Coverage Rates (%), and Two-Sigma Coverage Rates (%) for Substation and Feeder Loads (Case 6 in Example 1)

	MAPE	**MAE**	**STD**	**One-sigma Coverage**	**Two-Sigma Coverage**
Substation	5.17	2.83	4.46	81.07	96.70
Feeder 1	5.63	0.67	0.69	69.42	91.32
Feeder 2	5.73	0.54	0.61	69.18	92.26
Feeder 3	5.91	0.57	0.59	68.67	90.45
Feeder 4	7.25	0.45	0.62	75.78	94.10
Feeder 5	5.35	0.59	0.71	70.05	94.28
Feeder 6	5.44	0.38	0.45	73.78	92.30

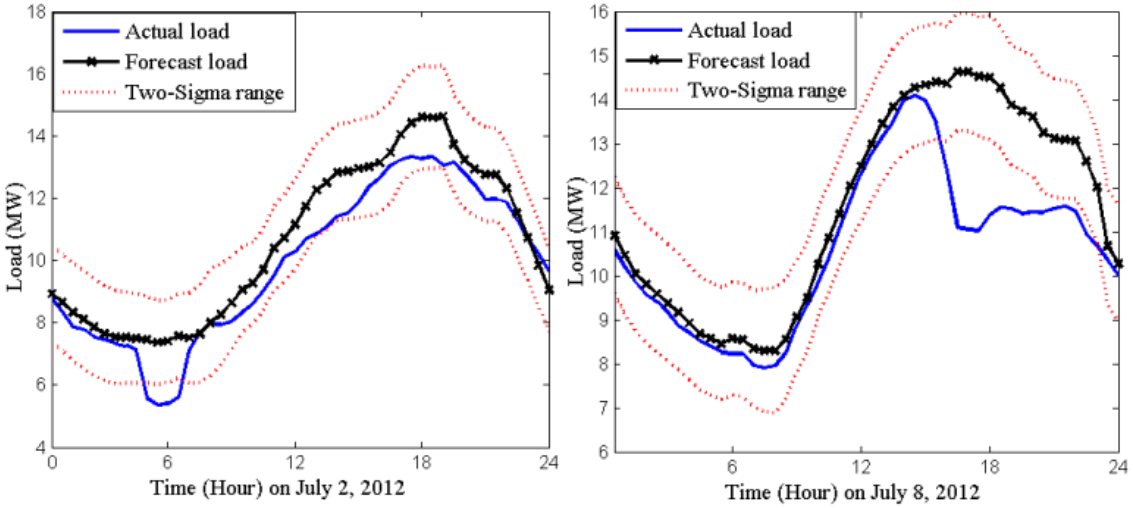


Figure 2.12. The switching operations detecetd for Feeder 1 on July 2 and July 8.

Case 7: This case compares our method with two naïve benchmarks, two multiple linear regression (MLR) models, and one back propagation NN (BPNN). The naïve benchmarks include using previous day load (Day-1) and the load of a week ago (Day-7) as forecasts for the next day load. The first MLR model has the corresponding independent variables as input factor of WNN as expressed in (2.13):

$$L(t) = \beta_0 + \beta_1 L_{SD}(t) + \beta_2 L(t-48) + \beta_3 T(t) + \beta_4 H(t) + \beta_5 I_{Day}(t) + \beta_6 I_{Hour}(t) + \varepsilon(t). \qquad (2.13)$$

- $L(t)$ denotes the load at time instance t. L_{SD} (t) denotes the modified similar-day load at time instance t. $L(t\text{-}48)$ denotes the load of previous day at the same time instance for 30-minute data.
- $T(t)$ and $H(t)$ denote predicted wind chill temperature and humidex at time instance t, respectively.
- $I_{Day}(t)$ is the day index: 1 for Monday, 2 for Tuesday-Wednesday-Thursday, 3 for Friday, 4 for Saturday, and 5 for Sunday.
- $I_{Hour}(t)$ is the time instance index, ranging from 1 to 48 for 30-minute data.
- β_i (i = 0, 1, .., 6) is the coefficient of each independent variable.

- $\varepsilon(t)$ is i.i.d Gaussian random variable with zero mean and finite variance.

The second MLR model has the same formulation but does not contain the modified similar day load $L_{SD}(t)$ variable. A traditional BPNN (implemented by MATLAB Neural Network Toolbox) without wavelet decomposition is also considered for comparison. Input factors for BPNN are identical with those used in WNN.

Forecasting MAPEs for the substation and six feeders are summarized in Table 2.10. As can be seen, benchmarks using previous day load and the load of a week ago encounter high prediction errors. The MLR without considering the similar day load also results in high MAPEs. However, when modified similar day load is taken into account, the prediction power of MLR is enhanced significantly. The MAPEs obtained from a single BPNN further outperforms MLRs because the nonlinear representation of input-output relationship in NN can better capture the load characteristics than the linear relationship used in MLR. Overall, our method achieves the lowest MAPEs for both substation and feeders because it captures the complicated load features through wavelet decomposition and dynamic pattern similarity analysis. Moreover, when compared with the forecasting MAPEs ranging from 2% to 20% for substation and feeder loads reported in the literature [24], [25], our results are competitive.

The NNs are trained offline by the historical data. Averagely, the training can be conducted within 20 minutes for both WNN and BPNN. Training times for WNN and BPNN are similar because even though BP iteration is faster than the decoupled EKF, BP converges much slower and requires a larger number of iterations. Once the WNN model is well trained and validated, it can be used to predict the next day load given the new input data. The weights of WNNs are then updated online with the realized latest 24 hours' loads.

TABLE 2.10 MAPE (%) for Substation July, 2012 Load (Case 7 in Example 1)

	A	D-1	D-7	MLR (1)	MLR (2)	BPNN
Substation	5.17	9.06	14.57	5.92	9.31	5.69
Feeder 1	5.63	7.13	11.12	6.08	9.70	5.97
Feeder 2	5.74	10.85	16.93	6.84	10.70	6.84
Feeder 3	5.91	10.19	16.85	6.67	10.21	6.35
Feeder 4	7.25	12.44	20.71	8.92	12.09	7.92
Feeder 5	5.35	8.82	13.31	6.75	9.31	6.51
Feeder 6	5.44	8.24	12.32	6.34	8.80	6.13

A: Using our method (WNN + LDF);
D-1: Using previous day load;
D-7: Using the load of a week ago;
MLR (1): Using MLR with selected similar day load;
MLR (2): Using MLR without selected similar day load;
BPNN: Using back propagation NN.

2.5.2 Example 2

This example demonstrates load forecasting for four feeders and selected customers in a city in Washington, United States. Load and weather data from December 01, 2011 to November 30, 2012 are collected. The training period is from December 2011 to September 2012, the validation period is October 2012, and the test period is November 2012. In addition, data of more than 1000 smart meters within this substation are available for four days from April 1, 2012 (a Sunday) to April 4, 2012.

Two cases are presented below. Case 1 predicts substation and feeder loads. Case 2 extends the forecasting to the customer level. In both cases, inputs for WNNs include similar day load, previous day load, weather predictions on the forecast day (wind-chill temperature and humidex), and day of week index.

Case 1: This case predicts the load of substation and demonstrates the values of switching operation detection. Among four feeders, Feeders 1 and 2 are forecast by LDF while Feeders 3 and 4 are forecast by WNN. As shown in Figure 2.13, a significant switching operation was detected on November 20, 2012 for Feeder 1. The load magnitude after switching operation decreased and Feeder 1 maintained the regular pattern for a certain period. Therefore, Feeder 1 is forecast by original LDF for the first twenty test days.

After November 20, Feeder 1 is forecast by the modified LDF, which is calculated based on the new data after the switching operation. Forecasting results of our method, two naïve benchmarks, two MLRs, and a simple BPNN are summarized in Table 2.11. It can be seen that our method provides the lowest MAPEs for the substation and all feeders.

Case 2: Smart meter-based customer load is predicted in this case. Since the smart meter data are limited and no transformer data are available, predictions are generated by using LDFs. Hourly load of April 4, 2012 is forecast based on feeder load forecasts and LDFs calculated by smart meter readings of the previous two days (Sunday April 1 is ignored). Forecasting results of four selected customers are shown in Table 2.12. Because of the complicated customer behaviors and the limited data, the law of large number is not as effective as for the aggregated load. Thus, high MAPEs are obtained as expected.

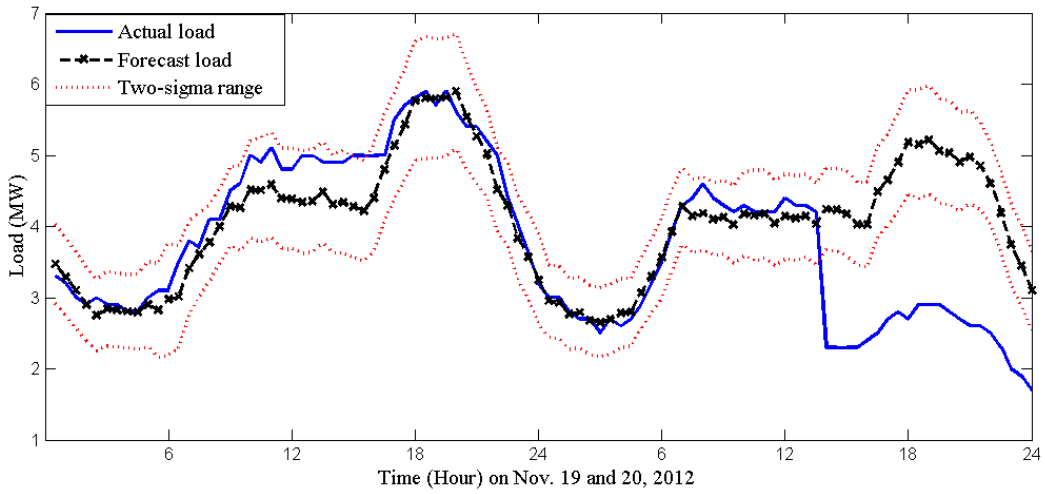


Figure 2.13. Switching operations detecetd for Feeder 1 on November 20, 2012.

TABLE 2.11
MAPE (%) FOR SUBSTATION AND FEEDER NOVEMBER, 2012 LOADS
(CASE 1 IN EXAMPLE 2)

	A	**D-1**	**D-7**	**MLR (1)**	**MLR (2)**	**BPNN**
Substation	3.80	7.89	12.45	6.12	8.66	5.03
Feeder 1	5.06	10.57	29.56	9.43	10.32	6.28
Feeder 2	4.18	8.94	13.46	6.24	9.31	5.71

Feeder 3	4.18	7.65	10.58	5.98	8.79	5.62
Feeder 4	3.47	6.47	7.63	6.04	8.36	4.94

A: Using our method (WNN + LDF);
D-1: Using previous day load;
D-7: Using the load of a week ago;
MLR (1): Using MLR with selected similar day load;
MLR (2): Using MLR without selected similar day load;
BPNN: Using back propagation NN.

TABLE 2.12
MAPE (%) FOR SMART METER NOVEMBER, 2012 LOAD
(CASE 2 IN EXAMPLE 2)

Smart Meter	**MAPE**
1	101.2
2	105.1
3	61.08
4	37.87

2.6 Conclusion

This Chapter presents a generic approach for short-term load forecasting at the distribution level within the hierarchical structure. With a root node forecast by WNNs and load pattern similarity-based child node classification, forecasts of all child nodes are obtained. Simple and fast LDF method is applied to forecast the load of regular child nodes, whereas WNNs with correlated sibling node are used for irregular nodes. The SPC-based switching operation detection is also considered to improve the forecasting accuracy. Testing results demonstrate the high prediction accuracy of this method. The new approach represents an effective way to forecast distribution-level load and would be helpful in the future smart grid.

References

[1] H. K. William, Distribution system modeling and analysis, 3rd ed., Florida: CRC Press, 2011.

[2] M. E. Baran, L. A. A. Freeman, F. Hanson and V. Ayers, "Load estimation for load monitoring at distribution substations," *IEEE Trans. Power Syst.*, vol. 20, no.1, pp. 164-170, Feb. 2005.

[3] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution system for loss reduction and load balancing," *IEEE Trans. Power Delivery*, vol. 4, pp. 1401-1407, Apr. 1989.

[4] M. S. Tsai and F. Y. Hsu, "Application of grey correlation analysis in evolutionary programming for distribution system feeder reconfiguration," *IEEE Trans. Power Syst*., vol. 25, no. 2, pp. 1126-1133, May 2010.

[5] H. L. Willis, *Spatial electric load forecasting*, 2nd edition, New York: Marcel Dekker, 2005.

[6] Electric load forecasting. [Online] Available: http://www.quanta-technology.com/sites/default/files/doc-files/Load_Forecasting-12-01-13.pdf.

[7] R. Sevlian and R. Rajagopal, "Value of aggregation in the smart grids," *IEEE Int. Conf. in Smart Grid Comm*., Oct. 2013.

[8] J. S. Taylor and N. Cristianini, Kernel methods for patter analysis, Cambridge University Press, June 2004.

[9] D. Sun, K. Cheung, K. Chung, and T. Mckeag, "System tools for integrating individual load forecasts into a composite load forecast to present a comprehensive synchronized and harmonized load forecast," U.S. Patent 0035071, Feb. 10, 2011.

[10] W. Guan, K. Chung, K. W. Cheung, X. Sun, L. D. Michel, S. Corbo and P. B. Luh, "Advanced load forecast with hierarchical forecasting capability," in *Proc. IEEE Power and Energy Society General Meeting*, Vancouver, BC, Canada, July. 2013.

[11] I. Moghram and S. Rahman, "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Transactions on Power Systems*, vol. 4, no. 4, pp. 1484-1491, Oct. 1989.

[12] Y. Bar-Shalom, X. R. Li and T. Kirubarajan, Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. New Jersey: Wiley, 2001.

[13] S. Haykin, Neural networks and learning machines, 3rd ed., NJ: Prentice Hall, 2009.

[14] A. Khotanzad, E. Zhou and H. Elragal, "A neuro-fuzzy approach to short-term load forecasting in a price-sensitive environment," *IEEE Trans. Power Syst*, vol. 17, no. 4, pp. 1273-1282, Nov. 2002.

[15] Y. Chen, P. B. Luh, C. Guan, Y. G. Zhao, L. D. Michel, M. A. Coolbeth, P. B. Friedland and S. J. Rourke, "Short-term load forecasting: similar day-based wavelet neural networks," *IEEE Transactions on Power Systems*, vol.25, no.1, pp. 322-330, Feb. 2010.

[16] S. Singhal and L. Wu, "Training feedforward networks with extended Kalman filter algorithm," in *Proc*. Int. Conf. ASSP, pp. 1187-1190, 1989.

[17] S. Haykin, Kalman Filtering and Neural Networks, NJ: Wiley, 2001.

[18] L. Zhang, P. B. Luh and K. Kasiviswanathan, "Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method," *IEEE Transactions on Power Systems*, vol. 120, no. 1, pp. 59-66, Feb., 2005.

[19] M. Espinoza, J. Suykens, R. Belmans, and B. D. Moor, "Electric load forecasting," *IEEE Control Syst. Mag.*, vol. 27, no. 5, pp. 43-57, Oct. 2007.

[20] S. Fan and Y. Wu, "Multiregion load forecasting for system with large geographical," *IEEE Transactions on Industry Applications*, vol. 45, no. 4, July/Aug. 2009.

[21] S. Fan, K. Methaprayoon and W. J. Lee, "Short-term multi-region load forecasting based on weather and load diversity analysis," *39th North American Power Symposium*, pp.562-567, 2007.

[22] N. Amjady, "Short-term bus load forecasting of power systems by a new hybrid method", *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 333-341, Feb. 2007.

[23] M. Espinoza, C. Joye, R. Belmans and B. De Moor, "Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1622-1630, Aug. 2005.

[24] Y. Goude, R. Nedellec and N. Kong, "Local short and middle term electricity load forecasting with semi-parametric additive models," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 440-446, Jan. 2014.

[25] K. N. Filho, A. D. P. Lotufo and C. R. Minussi, "Short-Term multimodal load forecasting using a general regression neural network," *IEEE Trans. Power Delivery*, vol. 26, no. 4, pp. 2862-2869, Oct., 2011.

[26] J. Yasuoka, J. L. P. Brittes, H. P. Schmidt and J. A. Jardini, "Artificial neural network-based distribution substation and feeder load forecast," *16th Int. Conf. Elect. Distribution*, 2001.

[27] X. Chen, C. Kang, X. Tong, Q. Xia and J. Yang, "Improving the accuracy of bus load forecasting by a two-stage bad data identification method," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1634-1641, Jul. 2014.

[28] B. Sun, P. B. Luh, Q. S. Jia, Z. O' Neill and F. Song, "Building energy doctors: an SPC and Kalman filter-based method for system-level fault detection in HVAC systems," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 1, pp. 215-229, Jan. 2014.

[29] C. Guan, P. B. Luh, L. D. Michel, Y. Wang, and P. B. Friedland, "Very short-term load forecasting: Wavelet neural networks with data prefiltering," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 30-41, Feb.

[30] R. R. Agnaldo and P. A. Alexandre, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 189–198, Feb. 2005.

[31] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," J. Energy, vol. 34, no. 1, pp. 46–57, Jan. 2009.

[32] G. Strang and T. Nguyen, Wavelets and Filter Banks, 2nd ed. Wellesley, MA: Wellesley-Cambridge Press, 1997.

[33] R. Gutierrez-Osuna, Lecture 13: Validation. [Online] Available: http://research.cs.tamu.edu/prism/lectures/iss/iss_l13.pdf

[34] A. Antoniadis, E. Paparoditis and T. Sapatinas, "A functional wavelet-kernel approach for time series prediction," *J. Royal Stat. Soc. B*, vol. 68, pp. 837–857, Nov. 2006.

[35] E. Keogh, K. Chakrabarti, M. Pazzani and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and Information Systems*, 2001.

[36] P. Esling and C. Agon, *Time-series data mining.* ACM Computing Surveys, Dec. 2012.

[37] C. Gu, D. Yang, P. Jirutitijaroen, W. M. Walsh and T. Reindl, "Spatial load forecasting with communication failure using time-forward kriging," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2875-2882, Nov. 2014.

[38] D. A. Bechrakis and P. D. Sparis, "Correlation of wind speed between neighboring measuring stations," *IEEE Trans. Energy Convers.*, vol. 19, no. 2, pp. 400-406, Jun. 2004.

[39] S. Pelland, G. Galanis and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model," *Progress in Photovoltaics: Research and Applications*, vol. 21, no.3, pp 284-296, May 2013.

[40] B. M. Hodge, D. Lew and M. Milligan, “Short-Term load forecasting error distributions and implications for renewable integration studies,” in *Proc*. 2013 *IEEE Green Technologies Conf.*, Denver, Colorado, April.

Chapter 3

A Novel Decomposition and Coordination Approach for Large Day-Ahead Unit Commitment with Combined Cycle Units

Day-Ahead Unit Commitment (UC) is an important problem faced by Independent System Operators (ISOs). Midcontinent ISO as the largest ISO in US, solves a complicated UC problem involving over 45,000 buses and 1,400 generation resources. With the increasing number of combined cycle units (CCs) represented by configuration-based modeling, solving the problem becomes more challenging. The state-of-the-practice branch-and-cut method suffers from poor performance when there are a large number of CCs. The goal of this Chapter is to solve such large UC problems with near-optimal solutions within time limits. In this Chapter, our recently developed Surrogate Lagrangian Relaxation, which overcomes major difficulties of Lagrangian Relaxation by not requiring dual optimal costs, is significantly enhanced through adding quadratic penalties on constraint violations to accelerate convergence. Quadratic penalty terms are linearized through a novel use of absolute value functions. Therefore, resource-level subproblems can be formulated and solved by branch-and-cut. Complicated constraints within a CC unit are thus handled within a subproblem. Subproblem solutions are then effectively coordinated. Computational improvements on key aspects are also incorporated to fine tune the algorithm. As demonstrated by MISO cases, the method provides near-optimal solutions within a time limit, and significantly outperforms branch-and-cut.

3.1 Introduction

Day-ahead unit Commitment (UC) is an important problem faced by Independent System Operators (ISOs). The UC problem is formulated as a Mixed Integer Linear Programming (MILP) problem, and has specified solving time limits and solution quality requirements. Midcontinent Independent System Operator (MISO), as the largest ISO in US, manages a network with over 45,000 buses and 1,400 internal generation resources. It also has a large number of (over 10,000) virtual trading undertaken by participants [1] in day-ahead market. Considering the extended network size and the increasing number of generation resources, solving the UC problem becomes challenging.

Combined Cycle units (CCs), as in its name, combine Combustion Turbines (CTs) and Steam Turbines (ST) into one unit. Since high-temperature gas from CTs is not released into the atmosphere but is used by STs to generate extra power, a CC unit produces electricity at high efficiency and with low CO_2 emissions. There is thus an upward trend of installing CC units worldwide [2]. However, CC units bring significant challenges to UC problems in view of their complicated operation characteristics. A CC unit can operate in different modes or configurations, each with a particular set of commitment states of CTs and STs. Switching commitment states among CTs and STs should follow pre-defined configurations and allowable transition paths. For example, a ST cannot generate electricity if there is not enough heat from CTs. In view of the complicated transitions among configurations, a CC unit is commonly modeled as an aggregated unit [3], which does not utilize possible transitions. Some ISOs, such as MISO and CAISO, are looking into configuration-based modeling to save energy costs and incentivize participants to offer true marginal costs [4]-[7].

The current state-of-the-practice to solve large UC problems is using commercial MILP solvers, which are Branch-and-Cut (B&C) based and combined with heuristics. The B&C method attempts to obtain a convex hull of the entire problem by using valid cuts that gradually cut off areas outside the convex hull. The optimal solution can then be obtained at one of vertices by solving a linear programming

problem. The B&C method explores problem linearity, but has no "local" concept. With the aggregated CC modeling, B&C generally performs well. However, when there are many CCs with configuration-based modeling, configuration transition constraints within a CC unit complicate the entire solution process. Significant computational efforts are required since cuts generated by B&C thus may not be tight and heuristics may not be effective. For a particular MISO case with 1,092 conventional units, 80 CCs (as stress tests), and about 15,000 virtuals looking ahead 36 hours, the method could not obtain a feasible solution with MIP gap less than 3% in 3,600s. The goal of this Chapter is to solve such large and difficult problems for near-optimal solutions within time limits using MILP solvers.

To solve such a large and difficult problem, a decomposition and coordination approach is preferred. Lagrangian Relaxation (LR) was traditionally used to solve UC to exploit problem separability [8]. However, standard LR suffers from major difficulties. Our recently developed Surrogate Lagrangian Relaxation (SLR) overcomes difficulties of LR by not requiring the optimal dual value and not fully optimizing the relaxed problem [9]. SLR is further combined with B&C to simultaneously exploit separability and linearity [10]. However, when there are a large number of CC units, levels of constraint violations may not be reduced sufficiently fast within a time limit.

In this Chapter, SLR is significantly enhanced by adding quadratic penalties on constraint violations to accelerate the convergence for large and difficult UC problems. Surrogate Augmented Lagrangian Relaxation (SALR) is thus developed with subproblem formulations and effective coordination. Due to the introduction of quadratic penalties, the augmented relaxed problem is nonlinear, and existing Mixed Integer Quadratic Programming (MIQP) solvers are not efficient to solve large-scale problems. To linearize quadratic penalties, a novel use of absolute value functions is established by exploiting the fact that an absolute value function and a quadratic function have the same minimum. This linearization approach is much more accurate than other traditional linearization methods. The relaxed problem is then decomposed into individual resource subproblems solved by B&C. Subproblem solutions are then

coordinated based on updating multipliers after solving one or multiple subproblems subject to surrogate optimality condition. For such a large problem, every aspect should be carefully handled to achieve an overall good performance. Several enhancements on key aspects including grouping resources in subproblem formulation, filtering out inactive transmission constraints, and developing effective heuristics to search feasible solutions are provided. Our work is timely and critical to solve large UC problems with increasing number of configuration-based CC units, and can be extended to other complicated large-scale MILP problems in power system and beyond.

The rest of the Chapter is organized as follows. Section 2.2 reviews the CC modeling and solution methodologies for UC problems with CC units. In Section 2.3, the UC formulation with configuration-based CC modeling is summarized. In Section 2.4, the solution methodology including linearization scheme, subproblem formulation, subproblem solution coordination, and convergence proof is presented. In Section 2.5, computational improvements on key aspects are introduced to enhance the overall performance. In Section 2.6, numerical testing results of one simple example and two MISO datasets are presented to demonstrate the efficiency and robustness of our method.

3.2 Literature Review

3.2.1 Combined Cycle Unit Modeling

A CC unit typically consists of one to four CTs and one or two STs, and may operate in different modes/configurations based on combinations of commitment states of CTs and STs. Considering the widely used linear solvers to solve UC problems, CC units are typically modeled within MILP framework in three different representations. The simplest one is the aggregated modeling, which represents a CC unit as a conventional unit, ignoring all possible transitions among different modes.

A second CC representation is the component-based modeling, which represents CTs and STs as individual components with its own unit parameters, such as ramping rate limits, minimum up/down time

limits, and startup/shutdown costs. Transitions within a CC unit are specified based on specific characteristics of the unit [11]. This model describes operating constraints for each component but requires modeling coupling steam constraints and may incur significant computational complexity especially when the number of CCs is large.

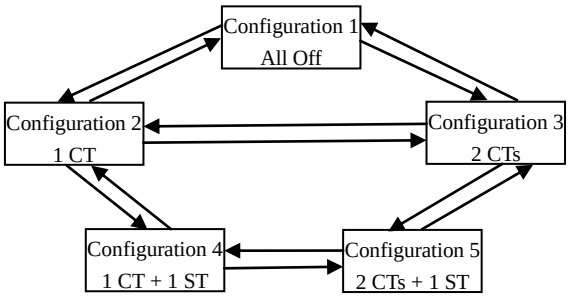


Figure 3.1. Allowable transitions in a CC unit with 2 CTs and 1 ST.

Another CC representation is the configuration-based modeling, which captures CC unit operational characteristics by using multiple configurations with possible transitions. Figure 3.1 shows allowable transitions among five configurations within a CC unit containing two CTs and one ST [12]. In this modeling [5], [10], [12] each configuration was modeled as a conventional unit and had its own operating characteristics. Constraints such as generation limits and ramping rates of a CC configuration were formulated following those of a conventional unit. There are several ways to model transitions. In [10], startup/shutdown variables and big "M" operations were used to indicate a transition among configurations. However, it led to major computational efforts. In [5] and [12], transition variables were used such that transition constraints can be directly captured in linear forms.

3.2.2 Methodologies for UC with CC Units

The B&C-based MILP solvers are widely used to explore problem linearity with the above CC presentations [5], [12]. The B&C method attempts to obtain a convex hull of the problem by using valid

cuts that gradually cut off areas outside the convex hull. The optimal solution can then be obtained at one of vertices by solving a linear programming problem. However, strongest possible cuts that define facets of the convex hull are problem-dependent and may not be easily obtained. In the absence of these facet-defining cuts, branching operations and heuristics are required. MILP solvers typically work well for UC problems with the aggregated CC modeling. However, when there are many CC units with configuration-based modeling, transitions among commitment states significantly complicate the problem convex hull. Cuts generated may not be effective, and time-consuming branching operation or heuristics are typically required.

Lagrangian Relaxation (LR) with subgradient methods to update multipliers was traditionally used to solve UC problems without CC units by exploiting separability [8]. After relaxing system-coupling constraints and decomposing the relaxed problem into subproblems, subproblem solutions are coordinated based on subgradient directions, which are obtained after solving all subproblems with given multipliers. However, standard LR requires the knowledge of the optimal dual value or its estimates.

Major difficulties of LR has been overcome by our recently developed surrogate Lagrangian relaxation (SLR) [9], in which surrogate subgradient directions are obtained after solving one or a few subproblems at a time subject to the simple surrogate optimality condition to ensure surrogate subgradient directions form acute angles with the direction toward optimal multipliers. Computational effort and multiplier zigzagging are much reduced. More importantly, convergence to the optimal multipliers has been proved based on contraction mapping and does not require the knowledge of the optimal dual value. Reference [9] is mainly on the theory with no large and difficult MILP problems tested.

The SLR has been further synergistically combined with B&C to simultaneously exploit separability and linearity. The SLR+B&C had been applied to a particular UC example with 300 conventional and 40 CC units – the latter modeled by using logical expressions [10]. Transmission capacity constraints are ignored, and each unit has a single-block cost function. Computational time was significantly reduced

compared with that of B&C. However, for MISO's cases with many transmission constraints, conventional and CC units with multiple-block cost functions, and virtual variables, levels of constraint violations are not reduced sufficiently fast to obtain near-optimal solutions within a limited solving time.

Augmented Lagrangian Relaxation (ALR) has a fast convergence by penalizing violations of coupling constraints [13]. Due to the introduction of quadratic terms, the relaxed problem becomes nonlinear and non-separable. The presence of quadratic terms can cause significant computational difficulties. Performance of existing MIQP solvers are often dramatically worse than that of linear cases. To address this issue, traditional way of linearization based on Taylor series expansion [14] was widely used. However, the approach is not effective since solution values tend to jump from one vertex to another as slopes of linear functions change. Adding proximal terms can alleviate this issue, and the proximal terms can be linearized following the way provided in PySP [15] by a simple interpolation. Our conference paper [16] summarized our working progress on linearization of ALR. It presented our investigation on maintaining the fast convergence of ALR by rewriting the entire Lagrangian function as summation of square functions with respect to each variable. However, due to the mathematical difficulties in completing squares and proof of preserving minima for practical problems as well as the implementation complexity, the method proposed in [16] is not applicable for large and practical UC problems.

3.3 Problem Formulation

Subsection 3.3.1 summarizes the UC problem formulation with constraints of configuration-based CC modeling descried in Subsection 3.3.2.

3.3.1 Unit Commitment Formulation

The goal is to commit and dispatch conventional units, CC units, as well as dispatchable variables including virtuals, dispatchable demands, and dispatchable transactions to minimize energy supply and reserve costs while satisfying individual resource-level constraints, power balance, reserve requirements,

and transmission capacity constraints for all looking ahead hours. The formulation presented in this section is based on MISO UC with configuration-based CC modeling [1], [12]. For CC units, each configuration is treated as a conventional unit with its own operating constraints. Transitions among configurations are developed by using transition variables and constraints. For presentation brevity, linear production costs are used and reserve is not considered. However, these features can be easily incorporated following [17], [18], and are considered in Examples 1-3 in Section 2.6.

For each conventional unit i at time t, decision variables include binary on/off status $x_{i,t}$, binary startup decision $u_{i,t}$, binary shutdown decision $y_{i,t}$, and continuous generation level $p_{i,t}$. For each configuration f of a CC unit j, decision variables include binary on/off status $x_{j,f,t}$, binary transition variable $v_{j,ff't}$ indicating a transition is made from configuration f to configuration f' following the allowable path, and continuous generation level $p_{j,f,t}$. Virtual, dispatchable demand, and dispatchable transaction variables $p_{n,t}$ are continuous and integer-independent.

Objective function

The objective of UC is to minimize total operating costs of all resources including conventional units (no-load cost $C_{NLi,t}x_{i,t}$, energy cost $C_{i,t}p_{i,t}$, and startup cost $C_{SUi,t}u_{i,t}$), CC units (no-load and energy costs for each configuration f, and transition cost $C_{j,f'f,t}v_{j,f'f,t}$ for each allowable transition) and virtuals, dispatchable demands, and dispatchable transactions (energy cost $C_{n,t}p_{n,t}$):

$$\min_{\{x,u,y,v,p\}} \sum_i \sum_t \left(C_{NLi,t}x_{i,t} + C_{i,t}p_{i,t} + C_{SUi,t}u_{i,t} \right) + \sum_j \sum_{f \in F_j} \sum_t \left(C_{NLj,f,t}x_{j,f,t} + C_{j,f,t}p_{j,f,t} + \sum_{f' \in F_j^{from,f}} C_{j,f'f,t}v_{j,f'f,t} \right) + \sum_n \sum_t C_{n,t}p_{n,t}, \tag{3.1}$$

where F_j denotes a set of configurations within CC unit j, and $F_j^{from,f}$ denotes a set of configurations f' ($f' \neq f$) that have an allowable transition to configuration f within CC unit j. Similarly, $F_j^{f,to}$ denotes a set of configurations f' ($f' \neq f$) that have an allowable transition from configuration f.

Individual conventional unit constraints

Conventional unit-level constraints include startup/shutdown, minimum up/down time, generator limits, and ramping up/down constraints [19].

Individual CC unit constraints

Constraints for a CC unit are summarized in Subsection 3.3.2.

Individual virtual transaction, dispatchable demand, and dispatchable interchange transaction constraints

Virtual trading is undertaken by participants that do not necessarily have physical loads to serve or physical resources to offer. Participants submit bids, either loads or supplies, for the financial purchase or sale of energy in the day-ahead market. Together with dispatchable demands and dispatchable transactions, these are associated with continuous variable $p_{n,t}$ only. Unlike generation resources, there is no integer variables and no other constraints except the MW limit constraint:

$$p_{n,t}^{\min} \leq p_{n,t} \leq p_{n,t}^{\max}, \forall n,t, \tag{3.2}$$

where $p_{n,t}^{\min}$ and $p_{n,t}^{\max}$ are limits with either $p_{n,t}^{\min} = 0$ or $p_{n,t}^{\max} = 0$. For a virtual supply offer, $p_{n,t}^{\min}$ is 0, and $p_{n,t}^{\max}$ is a positive value. For a virtual demand bid, $p_{n,t}^{\max}$ is 0, and $p_{n,t}^{\min}$ is a negative value. Dispatch demand

is always a negative value leading to $p_{n,t}^{\max} = 0$ and $p_{n,t}^{\min}$ as a negative limit. Dispatchable transactions could be either purchases or sales as similar to virtuals.

System-coupling power balance constraints

$$P_t = \sum_i p_{i,t} + \sum_j \sum_{f \in F_j} p_{j,f,t} + \sum_n p_{n,t}, \forall t, \tag{3.3}$$

where P is the sum of fixed demand bids at all nodes. At time t, P_t is equal to the total power generated by all generation resources and all other dispatchable variables.

System-coupling DC transmission capacity constraints

$$f_{l,t}^{\min} \le f_{l,t} \le f_{l,t}^{\max}, \ \ \forall l,t, \text{where } f_{l,t} = \sum_i \alpha_{i,l} p_{i,t} + \sum_j \sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t} + \sum_n \alpha_{n,l} p_{n,t} + P_{F_{l,t}}. \tag{3.4}$$

Power flow $f_{l,t}$ of transmission line l at time t is modeled as a linear function of injections from all nodes weighted by generation shift factors α plus a fixed demand $P_{F_{l,t}}$, and restricted by power flow limits.

3.3.2 Constraints of CC Units with Configuration-based Model

Within the configuration-based CC modeling, each configuration is treated as a conventional unit. Since $\sum_{f' \in F_j^{from,f}} v_{j,f',f,t}$ and $\sum_{f' \in F_j^{f,to}} v_{j,ff',t}$ can be viewed as startup and shutdown variable for configuration f, respectively, conventional unit-level constraints are applied to each configuration by using transition variables. The following are additional constraints defined to restrict configuration commitment states and transitions within a CC unit.

Configuration transition constraints

A CC unit can only have one configuration committed at *t*:

$$\sum_{f \in F_j} x_{j,f,t} = 1, \forall j, t. \tag{3.5}$$

Configuration startup/shutdown, minimum up/down time, and (3.5) guarantee $v_{j,f'f,t} = 1$ when there is a transition from configuration *f'* to *f*, and $v_{j,f'f,t} = 0$ otherwise [12].

Unique commitment constraints

These constraints have demonstrated strong ability in improving the computational performance [5] by defining that at most one transition is allowed at time *t* for a CC unit:

$$\sum_{f \in F_j} \sum_{f' \in F_j^{from,f}} v_{j,f'f,t} \leq 1, \forall j, t. \tag{3.6}$$

3.4 Solution Methodology

This section presents the solution methodology. Subsection 3.4.1 presents Surrogate Augmented Lagrangian Relaxation with a novel linearization scheme. Subsection 3.4.2 provides the convergence proof.

3.4.1 Surrogate Augmented Lagrangian Relaxation with absolute value function linearization

As reviewed in Subsection 3.2.2, our recently developed Surrogate Lagrangian Relaxation (SLR) overcomes major difficulties of LR. However, when there are many CC units, levels of violation of coupling constraints are still large when reaching solving time limits. This brings difficulty in searching feasible solutions. In addition, lower bounds may not provide a sufficiently good measure of solutions quality.

Motivated by the fast convergence of Augmented Lagrangian Relaxation, a novel approach Surrogate Augmented Lagrangian Relaxation (SALR) is developed by incorporating the idea of using quadratic penalties into SLR. Subproblems can then be extracted by fixing variables in other subproblems at values obtained from the previous iteration. Subproblem solutions are coordinated through updating multipliers subject to surrogate optimality condition. In the end, a near-optimal solution is obtained using heuristics.

After relaxing system-coupling constraints and penalizing violations, the augmented relaxed problem of UC at iteration *k* becomes:

$$\min_{\{x,u,y,v,p,w\}} L: \text{ with}$$

$$\begin{aligned} L \equiv & \sum_i \sum_t \left(C_{NLi,t} x_{i,t}^k + C_{i,t} p_{i,t}^k + C_{SUi,t} u_{i,t}^k \right) + \sum_n \sum_t C_{n,t} p_{n,t}^k + \sum_j \sum_{f \in F_j} \sum_t \left(C_{NLj,f,t} x_{j,f,t} + C_{j,f,t} p_{j,f,t} + \sum_{f' \in F_j^{from,f}} C_{j,f'f,t} v_{j,f'f,t} \right) \\ & + \sum_t \lambda_t^k \left(P_t - \sum_i p_{i,t}^k - \sum_j \sum_{f \in F_j} p_{j,f,t}^k - \sum_n p_{n,t}^k \right) \\ & + \sum_l \sum_t \mu_{l,t}^k \left(\sum_i \alpha_{i,l} p_{i,t}^k + \sum_j \sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t}^k + \sum_n \alpha_{n,l} p_{n,t}^k + P_{F_{l,t}} - f_{l,t}^{\max} + w_{l,t}^k \right) \\ & + \frac{c^k}{2} \sum_t \left(P_t - \sum_i p_{i,t}^k - \sum_j \sum_{f \in F_j} p_{j,f,t}^k - \sum_n p_{n,t}^k \right)^2 \\ & + \frac{c^k}{2} \sum_l \sum_t \left(\sum_i \alpha_{i,l} p_{i,t}^k + \sum_j \sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t}^k + \sum_n \alpha_{n,l} p_{n,t}^k + P_{F_{l,t}} - f_{l,t}^{\max} + w_{l,t}^k \right)^2, \end{aligned} \tag{3.7}$$

subject to individual resource-level constraints. In (3.7), λ^k and μ^k are Lagrangian multipliers (left-hand side of transmission constraints (3.4) are ignored in the presentation), c^k (>0) is a penalty coefficient, and $\{w_{l,t}\}$ are non-negative slack variables converting inequality constraints to equalities.

The augmented relaxed problem is nonlinear and non-separable because of the introduction of quadratic penalties. The presence of quadratic terms can cause significant computational difficulties. For large-scale problems, performance of existing Mixed Integer Quadratic Programing (MIQP) solvers is dramatically worse than that of linear cases. To maintain the linearity so as to exploit MILP solvers, our idea is to linearize quadratic terms by a novel use of absolute value functions. Conceptually, a quadratic

function can be replaced by a V-shape absolute value function with the minimum preserved as shown in Figure 3.2 for a simple quadratic function $y = x^2$. Although not differentiable, absolute value functions have the advantage of being exactly linearizable through extra variables and constraints.

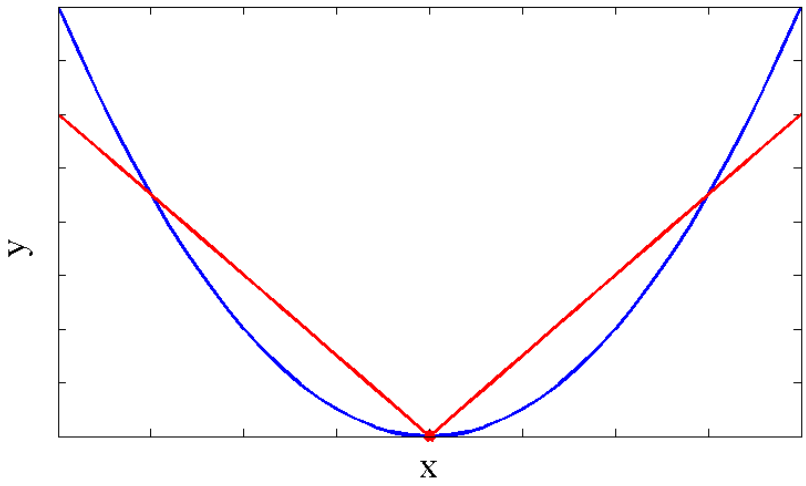


Figure 3.2. Absolute value function linearization.

Linearization

A simple MILP problem (8) is used to illustrate the linearization scheme.

$$\min_{x} \; f(x), \text{ s.t. } g(x) = 0, \text{ where } f \text{ and } g \text{ are linear.} \tag{3.8}$$

After relaxing the constraint, penalizing the violation, the augmented relaxed problem at iteration k becomes

$$\min_{x^k} \; f(x^k) + \lambda^k \cdot g(x^k) + 0.5c^k \cdot g(x^k)^2. \tag{3.9}$$

During the iterative process, replace $g(x^k)^2$ with an absolute value function:

$$\min_{x^k} \; f(x^k) + \lambda^k \cdot g(x^k) + 0.5c^k \cdot a^k \cdot \left| g(x^k) \right|, \tag{3.10}$$

where a^k is the slope of the absolute value function determined by the value of variable obtained from the previous iteration. At the beginning iterations, $a^k = \max\left(\left|g(x^{k-1})\right|, 1\right)$ is used in case $\left|g(x^{k-1})\right|$ becomes zero, and the constraint violation is not penalized. After a few iterations, the slope is fixed at a certain constant $\bar{a}$ to ensure the convergence.

The absolute value function is then linearized following a standard procedure [20] by introducing two non-negative variables (z_1, z_2) and one additional constraint. Equation (3.10) can be equivalently rewritten in a linear form as:

$$\min_{\{x^k, z_1^k, z_2^k\}} f(x^k) + \lambda^k \cdot g(x^k) + 0.5c^k \cdot a^k \cdot \left(z_1^k + z_2^k\right) \tag{3.11}$$

$$\text{s.t.} \, g(x^k) = z_1^k - z_2^k; \; z_1^k \geq 0; \; z_2^k \geq 0. \tag{3.12}$$

Following the above linearization scheme, the augmented relaxed problem at iteration *k* for UC can be linearized as:

$$\begin{aligned}
&\min_{\{x,u,y,v,p,w,z\}} L' : \text{with} \\
&L' \equiv \sum_i \sum_t \left(C_{NLi,t} x_{i,t}^k + C_{i,t} p_{i,t}^k + C_{SUi,t} u_{i,t}^k\right) + \sum_n \sum_t C_{n,t} p_{n,t}^k \\
&\quad + \sum_j \sum_{f \in F_j} \sum_t \left(C_{NLj,f,t} x_{j,f,t} + C_{j,f,t} p_{j,f,t} + \sum_{f' \in F_j^{from,f}} C_{j,f'f,t} v_{j,f'f,t} \right) \\
&\quad + \sum_t \lambda_t^k \left(P_t - \sum_i p_{i,t}^k - \sum_j \sum_{f \in F_j} p_{j,f,t}^k - \sum_n p_{n,t}^k \right) \\
&\quad + \sum_l \sum_t \mu_{l,t}^k \left(\sum_i \alpha_{i,l} p_{i,t}^k + \sum_j \sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t}^k + \sum_n \alpha_{n,l} p_{n,t}^k + P_{F_{l,t}} - f_{l,t}^{\max} + w_{l,t}^k \right) \\
&\quad + \frac{c^k}{2} \sum_t a_t^k \left(z_{1,t}^k + z_{2,t}^k\right) + \frac{c^k}{2} \sum_l \sum_t b_{l,t}^k \left(z_{3,l,t}^k + z_{4,l,t}^k\right),
\end{aligned} \tag{3.13}$$

subject to individual resource-level constraints and additional constraints to linearize absolute value functions:

$$z_{1,t}^k - z_{2,t}^k = P_t - \sum_i p_{i,t}^k - \sum_j \sum_{f \in F_j} p_{j,f,t}^k - \sum_n p_{n,t}^k, \forall t, \tag{3.14}$$

$$z_{3,l,t}^k - z_{4,l,t}^k = \sum_i \alpha_{i,l} p_{i,t}^k + \sum_j \sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t}^k + \sum_n \alpha_{n,l} p_{n,t}^k + P_{F_{l,t}} - f_{l,t}^{\max} + w_{l,t}^k, \forall l, t, \tag{3.15}$$

$$z_{1,t}^k, z_{2,t}^k, z_{3,l,t}^k, z_{4,l,t}^k \geq 0, \forall l, t. \tag{3.16}$$

In (3.13), a_t^k and $b_{l,t}^k$ are slopes of absolute value functions defined as

$$a_t^k = \max\left(\left| P_t - \sum_i p_{i,t}^{k-1} - \sum_j \sum_{f \in F_j} p_{j,f,t}^{k-1} - \sum_n p_{n,t}^{k-1} \right|, 1 \right), \forall t, \tag{3.17}$$

$$b_{l,t}^k = \max\left(\left| \sum_i \alpha_{i,l} p_{i,t}^{k-1} + \sum_j \sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t}^{k-1} + \sum_n \alpha_{n,l} p_{n,t}^{k-1} + P_{F_{l,t}} - f_{l,t}^{\max} + w_{l,t}^{k-1} \right|, 1 \right), \forall l, t. \tag{3.18}$$

Decomposition

Eq. (3.13) is now linear, and variables coupled in the original quadratic terms are now coupled through (3.14) and (3.15). Subproblems can still be extracted by fixing decision variables of other resources at values obtained from the previous iteration. The augmented relaxed problem is thus "decomposed" into individual resource-level subproblems. A subproblem for CC unit *j* at iteration *k* is formulated as:

$$\min_{\{x,v,p,w,z\}} L'_{CC} : \text{with } L'_{CC} \equiv \sum_{f \in F_j} \sum_t \left(C_{NL\,j,f,t} x_{j,f,t}^k + C_{j,f,t} p_{j,f,t}^k + \sum_{f' \in F_j^{from,f}} C_{j,f',f,t} v_{j,f',f,t}^k \right) + \sum_t \lambda_t^k \left(- \sum_{f \in F_j} p_{j,f,t}^k \right)$$

$$+ \sum_l \sum_t \mu_{l,t}^k \left(\sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t}^k + w_{l,t}^k \right) + \frac{c^k}{2} a^k \sum_t \left(z_{1,t}^k + z_{2,t}^k \right) + \frac{c^k}{2} b^k \sum_l \sum_t \left(z_{3,l,t}^k + z_{4,l,t}^k \right), \tag{3.19}$$

subject to individual CC unit constraints and constraints to linearize absolute value functions:

$$z_{1,t}^k - z_{2,t}^k = P_t - \sum_i p_{i,t}^{k-1} - \sum_{j' \neq j} \sum_{f \in F_j} p_{j,f,t}^{k-1} - \sum_n p_{n,t}^{k-1} - \sum_{f \in F_j} p_{j,f,t}^k, \forall t, \tag{3.20}$$

$$z^k_{3,l,t} - z^k_{4,l,t} = \sum_i \alpha_{i,l} p^{k-1}_{i,t} + \sum_{j' \neq j} \sum_{f \in F_j} \alpha_{j,f,l} p^{k-1}_{j',f,t} + \sum_n \alpha_{n,l} p^{k-1}_{n,t} + \sum_{f \in F_j} \alpha_{j,f,l} p^k_{j,f,t} + P_{F_{l,t}} - f^{\max}_{l,t} + w^k_{l,t}, \forall l,t, \quad (3.21)$$

$$z^k_{1,t}, z^k_{2,t}, z^k_{3,l,t}, z^k_{4,l,t} \geq 0, \forall l,t. \quad (3.22)$$

Complicated transitions within a CC unit are now handled within a CC unit subproblem, and do not complicate the entire solving process. Note that for a subproblem, when solving from one iteration to the next, only coefficients (λ^k, μ^k, c^k, a^k, b^k) and values of variables from other subproblems are changed.

A conventional unit subproblem can be similarly formulated. Other dispatchable resources including virtuals, dispatchable demands, and dispatchable transactions are integer-independent and only with simple bounds. Due to the simplicity, the original quadratic forms of these individual resource subproblems can be taken to derive analytical solutions based on unconstrained minima and bounds with slack variables fixed. Ideally, individual resource-level subproblems can be solved iteratively. For MISO UC with a large number of conventional units (more than 1,000), CC units (about 50), and virtuals (more than 10,000), slow convergence would be expected when solving individual resource subproblem. In the Subsection 3.5.1, a grouping strategy to create subproblems will be discussed to improve the convergence.

Coordination

To update multipliers, subproblem solutions are coordinated subject to the surrogate optimality condition to ensure that multiplier directions form acute angles with directions toward optimal multipliers:

$$\tilde{L}'_{c^k}(x^k, u^k, y^k, v^k, p^k, w^k, z^k, \lambda^k, \mu^k) < \tilde{L}'_{c^k}(x^{k-1}, u^{k-1}, y^{k-1}, v^{k-1}, p^{k-1}, w^{k-1}, z^{k-1}, \lambda^k, \mu^k), \quad (3.23)$$

where $\tilde{L}'_{c^k}(x^k, u^k, y^k, v^k, p^k, w^k, z^k, \lambda^k, \mu^k)$ is the surrogate augmented dual value of (3.13). After satisfying (3.23), multipliers are updated using stepsizes s^k as:

$$\lambda^{k+1}_t = \lambda^k_t + s^k \cdot \tilde{h}^k_t, \quad \mu^{k+1}_{l,t} = \mu^1_{l,t} + s^k \cdot \tilde{e}^k_{l,t}, \quad (3.24)$$

where $\tilde{h}_t^k$ and $\tilde{e}_{l,t}^k$ are components of surrogate subgradients $\tilde{g}(p^k, w^k)$. The stepsizing formula developed in SLR [9, (14)-(15), (20)] is used:

$$s^{k+1} = \gamma^k \frac{s^k \left\| \tilde{g}(p^{k-1}, w^{k-1}) \right\|}{\left\| \tilde{g}(p^k, w^k) \right\|}, \qquad (3.25)$$

where

$$\gamma^k = 1 - \frac{1}{Mk^d}, d = 1 - \frac{1}{k^r}, M \geq 1, 0 < r < 1 \quad (3.26)$$

Penalty coefficient is increased as:

$$c^{k+1} = \beta \cdot c^k, \ \beta > 1. \qquad (3.27)$$

If surrogate optimality condition (3.23) is not satisfied after solving all subproblems, penalty coefficients will be reduced to enforce the satisfaction of the surrogate optimality condition as:

$$c^{k+1} = c^k / \beta, \ \beta > 1. \qquad (3.28)$$

SALR with the absolute value function linearization is named SALRL for short. When synergistically combined with B&C to solve subproblems, SALRL+B&C is defined. The key steps of SALRL+B&C are summarized as follows:

Step 0 Initialize λ^0, μ^0, s^0, c^0, a^0, and b^0. Multipliers are initialized as 0.

Step 1 For given λ^k and μ^k, create a subproblem and solve it by B&C. If (3.23) is satisfied, go to next step. If the surrogate optimality condition is not satisfied, skip this iteration and solve the next subproblem. If the surrogate optimality condition is not satisfied after solving all subproblems, update penalty coefficients c^{k+1} per (3.28).

Step 2 Update γ^k and s^k per (3.25)-(3.26), update λ^{k+1} and μ^{k+1} per (3.24) and update c^{k+1} per (3.27).

Step 3 If a stopping criteria (a time limit is used for MISO UC problem) is satisfied, go to the next step. Otherwise, go back to Step 1.

Step 4 Search for feasible solutions.

Step 5 Calculate the duality gap by using the feasible solution and the dual value.

3.4.2 Convergence Proof

This subsection presents the convergence proof of SALR and SALRL+B&C methods.

Proposition 1: The SALR method converges if surrogate optimality condition

$$\tilde{L}_{c^k}(x^k,u^k,y^k,v^k,p^k,w^k,\lambda^k,\mu^k) < \tilde{L}_{c^k}(x^{k-1},u^{k-1},y^{k-1},v^{k-1},p^{k-1},w^{k-1},\lambda^k,\mu^k), \qquad (3.29)$$

in which $\tilde{L}_{c^k}(x^k,u^k,y^k,v^k,p^k,w^k,\lambda^k,\mu^k)$ is the surrogate augmented dual value of the augmented relaxed problem, is satisfied after solving all subproblems once, and penalty coefficient approaches a finite constant.

Proof: Within SALR, the augmented Lagrangian function (3.7) can be viewed as the Lagrangian function associated with Problem A:

$$\min_{\{x,u,y,v,p,w\}} f(x,u,y,v,p) + 0.5c \cdot g(p,w)^2, \qquad (3.30)$$

s.t. all system-coupling and resource-level constraints.

This problem is equivalent to the original UC problem when c approaches a finite constant ensuring no change in the objective function. Moreover, Problem A satisfies all assumptions listed in SLR (linearity of constraints [9, p. 178] and boundedness of constraint norms [9, p. 176]; but linearity of the objective function is not required). Therefore, the surrogate Lagrangian relaxation framework can be used whereby

multipliers and stepsizes are updated in the same way as within the SLR method subject to the surrogate optimality condition, and the convergence proof in Theorem 2.1 of SLR [9, pp. 180-186] can be applied to SALR. If the surrogate optimality condition is not satisfied after solving all subproblems, penalty coefficients will be reduced to enforce the satisfaction of the surrogate optimality condition and the updating of multipliers. Multipliers thus converge to λ^*, μ^* which maximize the dual function:

$$q_{c^k}(\lambda, \mu) \equiv \min_{\{x,u,y,v,p,w\}} L_{c^k}(x, u, y, v, p, w, \lambda, \mu), \tag{3.31}$$

corresponding to the augmented relaxed problem.

Proposition 2: SALRL+B&C converges if the surrogate optimality condition (3.23) is satisfied after solving all subproblems once, penalty coefficient approaches a finite constant, and the slopes of absolute value functions are fixed at constant values.

Proof: Within SALRL+B&C, the linearized surrogate augmented Lagrangian function (3.13) can be viewed as the Lagrangian function associated with Problem B defined as follows:

$$\min_{\{x,u,y,v,p,w\}} f(x, u, y, v, p) + 0.5c \cdot a \cdot |g(p, w)|, \tag{3.32}$$

s.t. all system-coupling and resource-level constraints. The objective of Problem 2 is the objective of the original UC problem plus the absolute value function terms.

At the beginning of the iterative process, slopes $a = \max\left(|g(p^{k-1}, w^{k-1})|, 1\right)$ would accelerate the reduction of $g(p,w)$. After several iterations, a would be fixed at $\bar{a}$, and the penalty coefficient c approaches $\bar{c}$, the problem (3.32) would be equivalent to the original problem. Moreover, (3.32) satisfies all assumptions listed in SLR. Therefore, following Proposition 1, convergence of SALRL+B&C can be guaranteed.

3.5 Computational Improvements

This section presents improvements on key aspects of the algorithm. It includes a grouping strategy to improve the convergence in Subsection 3.5.1, a method to filter out inactive transmission constraints in Subsection 3.5.2, and an effective heuristic to search feasible solutions in Subsection 3.5.3.

3.5.1 Grouping Resources within the Same Type

As discussed in Subsection 3.4.1, a subproblem can be extracted by fixing values of coupled resources in other subproblems in (14)-(15). Ideally, individual resource-level subproblems can be formulated and then solved iteratively. However, for large MISO UC problem, there are more than 1,000 unit-wise subproblems, and over 10,000 virtual subproblems. The enormous number of subproblems and less information in each small subproblem bring difficulty in satisfying the surrogate optimality condition and result in slow convergence. Take a conventional unit as an example, if all other resources are fixed and the penalty coefficient is large, absolute value terms become dominant. Solution then tends to be "feasible" to satisfy the current constraint violation rather than "optimal" to the original primal problem. There may not exist solutions that can satisfy the surrogate optimality condition. The wasted efforts on solving these individual unit-wise subproblems make the algorithm inefficient or does not converge within a time limit. Moreover, due to the inefficiency in most commercial optimization packages, the total time of solving 1,000 subproblems is typically much longer than that of solving 10 subproblems in which each contains 100 units.

In view of a large amount of resources in MISO system and the difficulty of satisfying surrogate optimality condition in solving individual resource subproblems, the relaxed problem is decomposed into limited number of subproblems. In each subproblem, certain resources within the same type are solved together to improve the convergence and reduce the computational time. Conventional units are divided

into ten subproblems, each containing about 120 units. Decision variables of these coupled units are now solved together within

$$z_{1,t}^{k} - z_{2,t}^{k} = P_t - \sum_{i=1}^{120} p_{i,t}^{k} - \sum_{i=121}^{I} p_{i,t}^{k-1} - \sum_{j} \sum_{f \in F_j} p_{j,f,t}^{k-1} - \sum_{n} p_{n,t}^{k-1}, \forall t. \tag{3.33}$$

Solving such a subproblem could have more chances to satisfy surrogate optimal condition because the solution $p_{i,t}^{k}$ for these units could likely move from the previous iteration $p_{i,t}^{k-1}$. Improvements on grouping resources within the same type will be demonstrated in Case 1 of Example 2 in Section 3.6.

In view of the complicated transitions within a CC unit, the number of CC units in one subproblem should be limited (10 is used for MISO UC cases) to avoid the potential increased complexity. Virtuals, dispatchable demands, and dispatchable transactions are only associated with continuous variables and simple MW limit constraints. In view of these features, all these variables are solved together in one subproblem.

3.5.2 Identification of Inactive Transmission Constraints

Within SALRL+B&C, since all transmission constraints are relaxed and violations are penalized, a large number of such constraints would result in a large number of multipliers and absolute value functions. To overcome these difficulties, inactive transmission constraints are identified and removed following [21]. An analytical estimate of the worst-case power flow through each line is obtained. If it is within the transmission capacity, the corresponding transmission constraint is removed. This method is extended to CCs by only considering the worst-case configuration (all components “on” when the generation shift factor is positive, and all components “off” when the generation shift factor is negative). Virtuals, dispatchable demands, and dispatchable transactions are treated similarly to conventional units since they have MW limits. This process reduces solving time for subproblems and computational time of surrogate subgradients and multipliers.

3.5.3 Obtaining Feasible Solutions

Solutions to subproblems are typically feasible with respect to each subproblem, but these solutions may not satisfy relaxed constraints. To obtain a feasible solution to the original problem, heuristics are used to free certain discrete variables to solve a smaller MILP problem by B&C.

Considering the complexity associated with CC units, binary variables of CC units are fixed at values obtained from the iterative process. Level of violation for each transmission constraint at each time interval can be calculated as:

$$g_{l,t} = \sum_i \alpha_{i,l} p_{i,t} + \sum_j \sum_{f \in F_j} \alpha_{j,f,l} p_{j,f,t} + \sum_n \alpha_{n,l} p_{n,t} + P_{F_{l,t}} - f_{l,t}^{\max}. \qquad (3.34)$$

For example, if $g_{l,t} > 0$, it indicates that extra power is generated. Commitment variables of online units with positive $\alpha_{i,l}$ and off-line units with negative $\alpha_{i,l}$ are set free and will be resolved to potentially reduce the power flow. Number of selected units and magnitude thresholds of $\alpha_{i,l}$ can be tuned based on testing cases. Additional out-of-money units are identified if their subproblem costs are positive based on MISO's heuristics [1] to ensure that sufficient units are resolved in the small MILP problem. Since levels of constraint violations are penalized, SALRL+B&C provides better commitment decisions, and fewer units are freed as compared with those from SLR+B&C as demonstrated in Section 3.6.

3.6 Numerical Testing

The method has been implemented on an Intel Core i7-6700K 4.0GHz 16 GB server with AIMMS 4.2 and CPLEX 12.6. Three examples are presented. In Example 1, a 5-bus system is tested to demonstrate the efficiency of the configuration-based CC modeling and the convergence of SALRL+B&C. In examples 2 and 3, two different MISO datasets are tested to demonstrate near-optimal solutions are effectively obtained within a time limit.

3.6.1 Example 1: 5-Bus system

A 5-bus system with eight conventional units, one combined cycle unit, and 224 virtuals looking ahead 36 hours is tested. After relaxing system coupling constraints, the relaxed problems are decomposed into nine unit subproblems and one virtual subproblem. Iterative process stops when the number of iterations (solving one subproblem is defined as one iteration) reaches 200, feasible solution search is then started based on heuristics. The following two cases are studied to demonstrate the efficiency of configuration-based CC modeling and the convergence of our solution methodology.

Case1*:* This case compares the performance of configuration-based CC modeling and the aggregated CC modeling. In the former, the CC unit contains two CTs and one ST, and allows five configurations as shown in Figure 3.1. The feasible solution of configuration-based CC modeling is 0.54% less than that of the aggregated modeling. Comparisons of commitment states and generation levels are shown in Figure 3.3. As can be seen, in the aggregated modeling, the CC unit is committed in the last eight intervals with a high generation level. In the configuration-based modeling, the CC unit is committed four hours earlier and the generation level is reduced by switching to configurations with smaller minimum generation requirement. This is because in the aggregated modeling, high capacity, high startup/energy cost, and long minimum run time are used to ensure that the operating costs can be recovered. The configuration-based model introduces flexibility because different operational configurations can be selected and the energy costs and minimum up/down time are more accurate for each configuration than the simplified aggregated model.

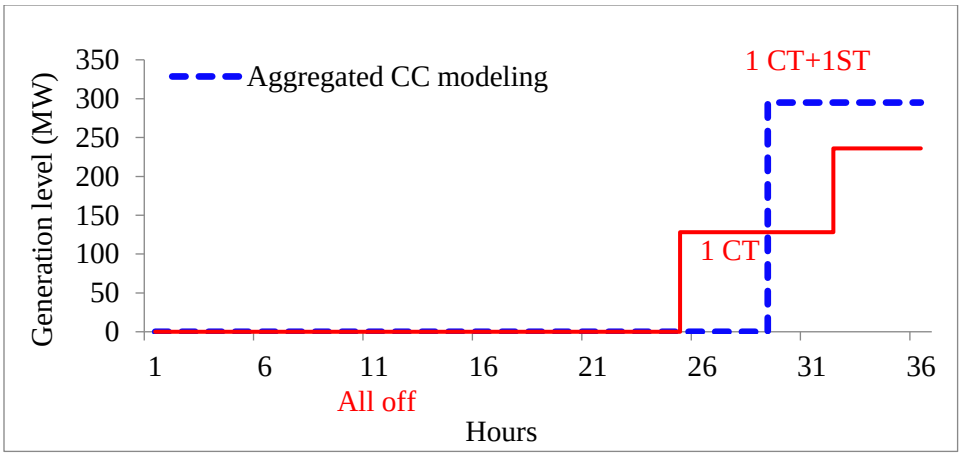


Figure 3.3. Generation level and configuration transitions of the CC unit.

Table 3.1. Performance comparisons of SALRL+B&C, SALR+MIQP, SLR+B&C, and ALR+MIQP.

	SALRL+ B&C	SALR + MIQP	SLR + B&C	ALR + MIQP
Feasible solution ($)	604,903	604,903	604,903	604,903
CPU time (s)	40	53	39	258

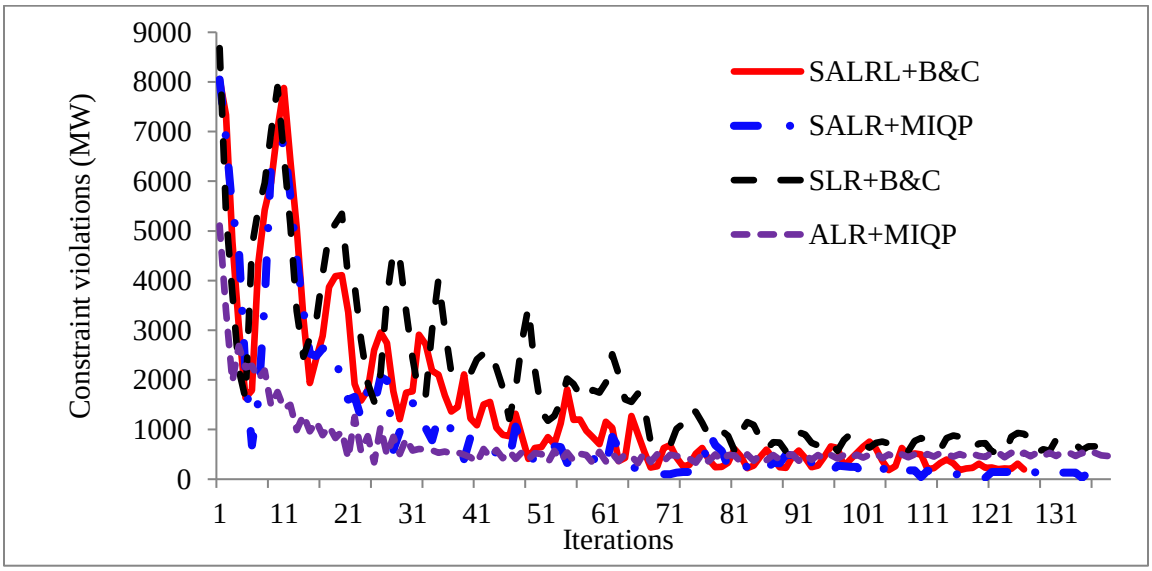


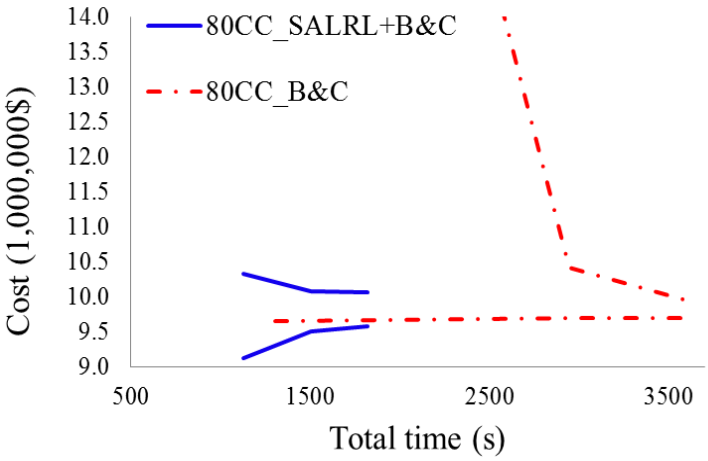
Figure 3.4. Convergence of SALRL+B&C compared with SALR+MIQP, SLR+B&C, and ALR+MIQP.

Case 2: This case compares the performance of SALRL+B&C, SALR+MIQP (with MIQP solver to solve subproblems), SLR+B&C, and ALR+MIQP (with MIQP solver to solve the full augmented relaxed problem). Feasible solutions and computational time of these four approaches are summarized in Table 3.1. All approaches obtain the same optimal solution $604,903. Among them, ALR+MIQP requires the longest time since the problem is more complex than other decomposition approaches during each iteration. The computational time of SALRL+B&C is similar to that of SLR+B&C demonstrating the efficiency of our linearization approach. SALR+MIQP requires longer computational time than SALRL+B&C because of the complexity to solve a quadratic problem. For this simple example with only a few units, MIQP works with acceptable performance. When solving large-scale MIQP problems, the complexity would increase exponentially.

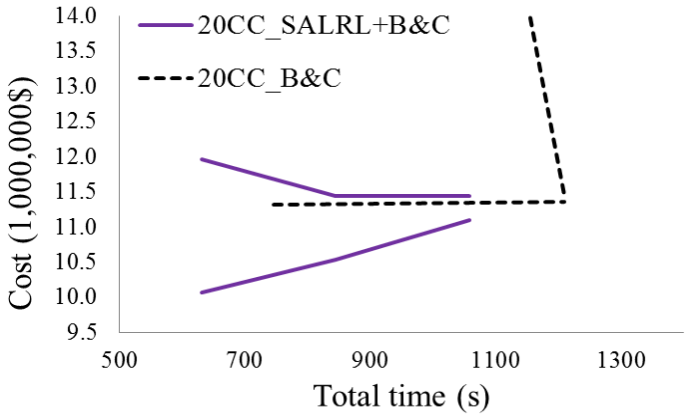
Levels of constraints violations in squared L2-norm are depicted in Figure 3.4 to illustrate the convergence of all approaches. The two SALR-based approaches accelerate the reduction of levels of constraint violations compared with SLR. Constraint violations of SALRL+B&C are slightly higher than that of SALR+MIQP because of the linear approximation. ALR approach reduces the constraint violation fast at the beginning because it fully optimizes the augmented relaxed problem. However, the level of constraint violations is still high at convergence due to the major issues as discussed in LR method.

3.6.2 Example 2

One MISO dataset containing 1,143 conventional units, 20 CC units converted from aggregated modeling to configuration-based modeling, and 14,955 virtuals with 36 looking ahead hours is tested. The case is also extended to 40 and 80 CC units with configuration-based modeling. All system-coupling constraints including power balance, transmission, and reserve are relaxed. Among them, power balance and transmission constraints are penalized with quadratic terms.



(a) 80 CCs case



(b) 20 CCs case

Figure 3.5. Comparison of SALRL+B&C and B&C for (a) 80 CCs, and (b) 20 CCs in Example 2.

Performance of SALRL+B&C and comparison with B&C are depicted in Figure 3.5. For 80 CCs as a stress test, at the time limit 1,800s, B&C cannot reach a feasible solution with a reasonable MIP gap. It requires 3,600s to first obtain a good feasible solution. The SALRL+B&C method provides a near-optimal solution with an acceptable duality gap 5.5% within 1,800s. Along the solution process, it achieves feasible solution much faster than B&C. Although the feasible objective cost obtained from B&C at 3,600s is slightly less than that of SALRL+B&C at 1,800s, the 3,600s solving time is too long for MISO to accept. For less complicated 20 CC case, B&C obtains a good feasible solution with a very

small MIP gap 0.8%, which is hard for SALRL+B&C to outperform. This demonstrates that SALRL+B&C is powerful to provide good-quality feasible solutions within limited time and significantly outperforms B&C for difficult cases.

The following cases 1-3 are conducted for 80 CCs to demonstrate the efficiency of our approach on grouping strategy, filtering out inactive transmission constraints, and searching for feasible solutions, respectively. In Case 4, our method is compared with B&C for 20, 40 and 80 CCs.

Case 1: This case shows the performance of grouping resources in subproblems. The nominal grouping strategy is defined as follows: 120 conventional units are grouped into one subproblem, 10 CCs are grouped into a CC unit subproblem, and all virtuals, dispatchable demands, and dispatchable transactions are grouped together into one subproblem. Therefore, totally 19 subproblems are created and solved iteratively. Comparisons of nominal grouping and solving individual unit-wise subproblems are presented in Table 3.2. Due to the large number of virtual variables, all virtuals and other dispatchable variables are solved together in both scenarios, and we only compare for generation units.

With the nominal grouping strategy, the initial model generation time and solving time for completing 19 subproblems once are 71s and 75s, respectively. The total time 164s also includes computational efforts on multiplier updating, data transfer, surrogate subgradient calculation, etc. When solving individual unit-wise subproblems, solving all 1,173 subproblems once requires more than 1,000s. The large computation time of solving individual unit-wise subproblems mainly results from the large overhead in AIMMS. Moreover, at the targeted 1,800s, levels of constraint violations with the nominal grouping strategy are small, and a near-optimal feasible solution is obtained. However, when solving individual unit-wise subproblems, since less information is used in each small subproblem, levels of constraint violations are not much reduced leading to a very slow convergence.

Table 3.2. Comparison of nominal grouping with solving individual unit-wise subproblem.

		Solving all subproblems once			At 1,800s
	# of sub-problems	Generation Time (s)	Solving Time (s)	Total Time (s)	MIP gap (%)
Nominal grouping	19	71	75	161	5.5
Individual unit-wise subproblem	1,173	312	392	1,042	>100

Case 2: This case compares the performance of SALRL+B&C with and without filtering inactive transmission constraints (TCs). Numbers of original and remaining TCs after filtering are shown in Table 3.3. Performance of SALRL+B&C with and without filering out inactive TCs are compared in Table 3.4. This process reduces computational time and improves the feasible solution.

Case 3: Performance of heuristics to search feasible solutions is shown in Table 3.5. The number of commitment variables selected based on generation shift factor is reduced by 21.9% from SLR, and the number of out-of-money units identified by SALRL+B&C is reduced from 136 to 112 as compared with SLR. Feasible solution is also improved.

Case 4: This case shows the scalability of SALRL+B&C with 20, 40 and 80 CC units as depicted in Figure 3.6. SALRL+B&C obtains near-optimal solutions for all scenarios within the targeted time. For 20 CCs and 40 CCs, SALRL+B&C provides good-quality feasible costs within 1,200s. When the number of CC units increases, the total cost is reduced because of the energy efficiency of configuration-based modeling. Meanwhile, computational time increases because more CC subproblems need to be solved and the feasible solution search becomes complicated. SALRL+B&C thus requires a longer time to obtain a small duality gap. As it can be seen, B&C is powerful to provide a good lower bound while our method provides a high-quality feasible solution. To take advantages of the two methods, we may run SALRL+B&C and B&C in a parallel manner. Along the iterative process, the lower bound from B&C can be taken and compared with our feasible solution to calculate the MIP gap.

Table 3.3. Numbers of original and remaining TCs.

Original TCs		Remaining TCs		Filtered
Right side	Left side	Right side	Left side	48.8%
8,347	7,032	5,503	2,372	

Table 3.4. Comparison of filtering and not filtering inactive TCs.

	UB* ($)	LB** ($)	Gap (%)	Total time (s)
SALRL+B&C[1]	10,064,124	9,525,597	5.5	1,896
SALRL+B&C[2]	10,082,321	9,517,789	5.6	1,995

[1]: w/ filtering inactive TCs; [2]: w/o filtering inactive TCs.
*: upper bound, the feasible solution.
**: lower bound.

Table 3.5. Comparison of UB search between SALRL and SLR.

	# of binary freed[1]	# of units freed[2]	Feasible cost ($) w/ 1	Feasible cost ($) w/ 1 and 2
SALRL+B&C	1,683	112	10,075,603	10,064,124
SLR+B&C	2,155	136	10,124,712	10,084,577

[1]: binary variables freed based on violations of transmission constraints;
[2]: number of out-of-money units freed based on MISO heuristic.

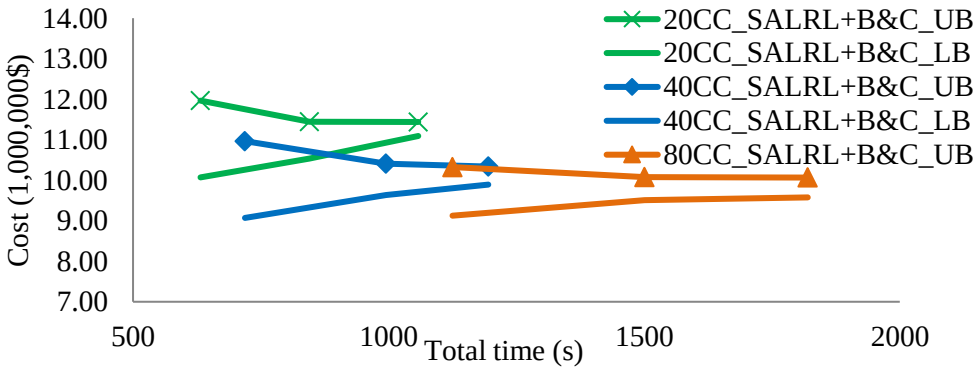
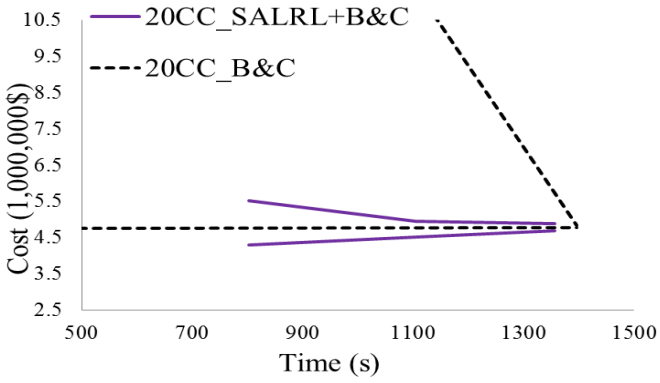


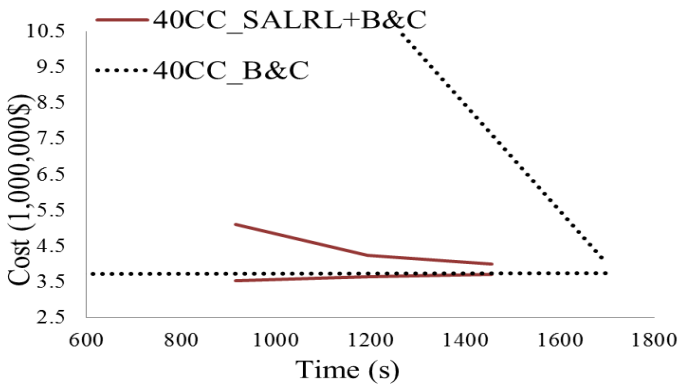
Figure 3.6. SALRL+B&C for 20, 40 and 80 CC units in Example 2.

Example 3

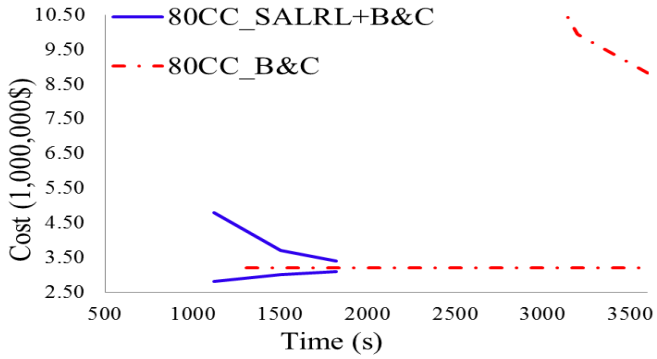
To test the robustness of our method, another MISO case containing 1,129 conventional units and 15,843 virtuals with 36 looking ahead hours is tested. Performance of SALRL+B&C for 20, 40, and 80 CC units is compared with B&C, SLR+B&C as in Fig 3.7. With 20 and 40 CC units, both SALRL+B&C and B&C methods obtain near-optimal solution within 1,800s. However, when the number of CC units increases to 80, the complexity increases exponentially and hits the bottom of B&C method. Therefore, B&C cannot obtain a feasible solution with a small MIP gap within limited time. Even after 3,600s, it is still difficult for B&C to find a feasible solution with an acceptable MIP gap. For 80 CC cases, SALRL+B&C is powerful to obtain a near-optimal solution at targeted 1,800s by fully exploiting exponential reduction of complexity. Our SALRL+B&C method also outperforms SLR+B&C in terms of lower bound, feasible solution, and computational time as shown in Table 3.6. It converges fast and reduces the duality gap along the iterative process. These improvements could result in significant annual savings for ISOs considering that such a large UC problem involves millions of dollars each day.



(a) 20 CCs case



(b) 40 CCs case



(c) 80 CCs case

Figure 3.7. Comparison of SALRL+B&C and B&C for (a) 20 CCs, (b) 40 CCs and (c) 80 CCs in Example 3.

Table 3.6. Results of SALRL+B&C, SLR, and B&C for 80 CCs in Example 3.

	UB ($)	LB ($)	Gap (%)	Total time (s)
SALRL+B&C	3,308,783	3,172,256	4.3	1843
SLR+B&C	3,329,845	3,151,576	5.6	1859
B&C	14,947,775	3,204,067	>100	1902

3.7 Conclusion

This Chapter targets to solve a large and difficult UC problem which contains over 1,000 units and 10,000 virtuals looking ahead 36 hours. With an increasing number of combined cycle units represented

by configuration-based modeling, current state-of-the-practice B&C cannot solve the problem with targeted MIP gap or within a time limit. Decomposition and coordination is a must for such large-scale complicated problems. Our recently developed surrogate Lagrangian relaxation is thus significantly enhanced through adding quadratic penalties on constraint violations to fully exploit exponential reduction of complexity with fast convergence. Quadratic terms are innovatively linearized through a novel use of absolute value functions. Enhancements on certain key aspects are also incorporated to fine tune the algorithm and improve the overall performance. As demonstrated by MISO cases, the method provides near-optimal solutions within a time limit, and significantly outperforms B&C.

Our work is timely and critical to solve large UC problems, and can be extended to other complicated MILP problems in power systems and beyond. There is still room to improve the method, and we propose two directions as in future work: 1) Formulation tightening and 2) Distributed and asynchronous implementation. Tightening subproblem formulation is important since if constraints directly delineate a problem convex hull, the MILP problem can be directly solved by linear programming. Distributed and asynchronous implementation of SALRL+B&C will also be investigated thus that subproblems will be solved in a distributed way and multipliers will be updated asynchronously to improve the efficiency.

References

[1] Y. Chen, A. Casto, F. Wang, Q. Wang, X. Wang, and J. Wan, "Improving large scale day-ahead security constrained unit commitment performance," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4732-4743, 2016.

[2] B. Lu and M. Shahidehpour, "Short-term scheduling of combined cycle units," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1616–1625, 2004.

[3] IESO, Enhanced Day-Ahead Commitment Combined-Cycle Modeling, Nov. 2008 [Online]. Available: http://www.ieso.ca/imoweb/ pubs/consult/se21-edac/se21-edac-20081128-Pseudo_Units.pdf

[4] M. Tamayo, X. Yu, X. Wang, and J. Zhang, "Configuration based combined cycle model in market resource commitment," in *Proc. 2013 IEEE Power and Energy Soc. General Meeting*, 2013.

[5] Y. Chen and F. Wang, "MIP formulation improvement for large scale security constrained unit commitment," *Electric Power systems Research*, vol. 148, pp. 147-154, 2017.

[6] CAISO, Multi-Stage Generator Unit Modeling Enhancements, Sep. 2011, Tech. Rep. [Online]. Available: http://www.caiso.com/Documents/DraftFinalProposal-Multi-StageGeneration Enhancements.pdf

[7] H. Hui, C.-N. Yu, F. Gao, and R. Surendran, "Combined cycle resource scheduling in ERCOT nodal market," in Proc. *2011 IEEE Power and Energy Soc. General Meeting*, 2011.

[8] S. Virmani, E. C. Adrian, K. Imhof, and S. Mukherjee, "Implementation of a Lagrangian relaxation based unit commitment problem," *IEEE Transactions on Power Systems*, vol. 4, no. 4, pp. 1373-1380, Nov 1989.

[9] M. A. Bragin, P. B. Luh, J. H. Yan, N. Yu, and G. A. Stern, "Convergence of the surrogate Lagrangian relaxation method," *Journal of Optimization Theory and Applications*, vol. 164, no. 1, pp. 173-201, 2015.

[10] M. A. Bragin, P. B. Luh, J. H. Yan, and G. A. Stern, "Surrogate Lagrangian relaxation and branch-and-cut for unit commitment with combined cycle units," *In Proceedings of the IEEE Power and Energy Society General Meeting*, National Harbor, Maryland, July 2014.

[11] C. Liu, M. Shahidehpour, Z. Li, and M. Fotuhi-Firuzabad, "Component and mode models for the short-term scheduling of combined-cycle units," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 976–990, 2009.

[12] G. Morales-España, C. M. Correa-Posada and A. Ramos, "Tight and compact MIP formulation of configuration-based combined-cycle units," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1350-1359, 2016.

[13] D. P. Bertsekas, Nonlinear Programming, 3rd ed. Belmont, MA, USA: Athena Scientific, 2016.

[14] C. Beltran and F. J. Heredia, "Unit commitment by augmented Lagrangian relaxation: Testing two decomposition approaches," *Journal of Optimization Theory and Applications*, vol. 112, no. 2, pp. 295–314, 2002.

[15] J. Watson, D. L. Woodruff, W. E. Hart, "PySP: Modeling and solving stochastic programs in Python," *Mathematical Programming Computation.*, vol. 4, no. 2, pp. 109-149, Jun. 2012.

[16] X. Sun, P. B. Luh, M. A. Bragin, Y. Chen, J. Wan, and F. Wang, "A decomposition and coordination approach for large-scale security constrained unit commitment problems with combined cycle units," *in Proc. 2017 IEEE Power and Energy Soc. General Meeting*, 2017.

[17] Alstom configuration-base combined cycle modeling Design Note, Alstom Grid, 2015.

[18] MISO e-terra Commit (RSC) Design Note (DN), Alstom Grid, 2015.

[19] G. Morales-Espana, J. M. Latorre, and A. Ramos, "Tight and compact MILP formulation for the thermal unit commitment problem," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4897–4908, Nov. 2013.

[20] J. Bisschop, AIMMS Optimization Modeling, *Lulu Enterprises Incorporated*, 2006.

[21] Q. Zhai, X. Guan, J. Cheng, and H. Wu, "Fast identification of inactive security constraints in SCUC problems," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1946-1954, 2010.

Chapter 4

Probabilistic Forecasting of Dynamic Thermal Ratings for Overhead Transmission Lines

Transfer capability of an overhead transmission line is limited by its thermal rating. The static thermal rating based on the worst-case weather condition does not utilize dynamic cooling and heating effects on conductors from ambient weather. The dynamic thermal rating (DTR) adapts the thermal capacity based on measured and predicted weather and typically results in a high rating without scarifying system security. Appropriate hourly DTR forecasts could be integrated with day-ahead market to yield economic benefits. Employing DTR forecasting, however, is difficult in view of many relevant weather factors involved in the thermal models, data availability, inherent weather uncertainties as well as the spatial topology of transmission lines. In this Chapter, a probabilistic DTR forecasting approach is established to resolve the above difficulties. Major weather factors are selected based on impact analysis, and a spatio-temporal regression model is developed for weather forecast with available weather sources. Spatial topology of transmission lines and weather uncertainties are captured by treating the line rating as a random variable, which is the minimum of selected span thermal capacities. The DTR distribution and the corresponding percentiles are then obtained with high accuracy and low computational requirements. Numerical testing results for a short transmission line segment and a long line demonstrate that our method provides secure and high ratings.

4.1 Introduction

Electric transmission lines are essential to deliver electricity to customers. The transfer capability is constrained by their thermal ratings, which are the highest current that lines can carry without damaging conductors. The normal thermal rating is the maximum loading that a conductor can carry continuously during steady state and is used in unit commitment, economic dispatch, and base case power flow analysis [1]-[3]. The emergency thermal ratings are the maximum loadings within specified periods, used for a fault duration or post-contingency analysis [1]-[4].

An overhead transmission line could be tens of kilometers long and consist of hundreds of spans in multiple line sections. As shown in Figure 4.1, a line segment between two adjacent towers is referred to a span, and a line section refers to a portion of the line that extends from one dead-end tower to another. Weather conditions such as ambient temperature and wind affect the thermal capacity of a span and result in different thermal capacities at different span locations throughout a day.

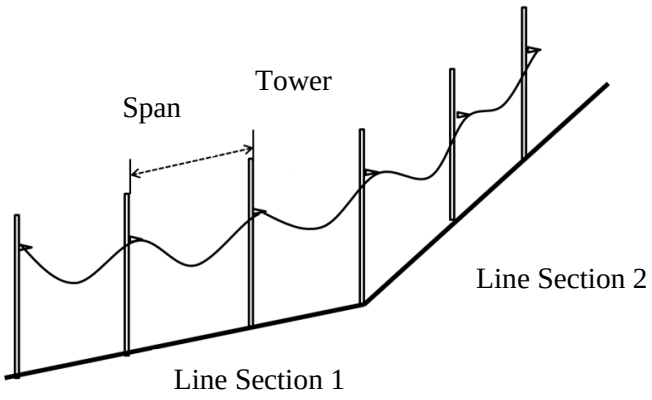


Figure 4.1. Layout of an overhead transmission line.

The static thermal rating (STR) is traditionally used to calculate the thermal capacity under the worst-case weather condition (e.g., 40°C summer ambient temperature, 0.6m/s perpendicular wind speed, and full sun) throughout a line. This conservative weather assumption may restrict the thermal capacity whenever the real ambient weather condition is less stressful. With high penetration of the renewable

energy, wind and solar farms built in remote areas that lack suitable power grid connections could result in system congestion. The STR, which does not consider weather and thermal capacity changes along the line and throughout the day, is thus not good enough.

The dynamic thermal rating (DTR) adapts the thermal capacity based on measured or predicted weather and typically results in a high rating without scarifying system security. If high accuracy hourly DTR forecasts are available, they could be integrated with day-ahead unit commitment and power flow analysis to yield economic benefits. It has been shown that the DTR helped reduce energy costs by 10% compared with the STR and relieved transmission congestions [5]. Inherent uncertainties in weather forecasts could result in uncertainties in DTR calculation. The estimated DTR may exceed the actual thermal capability of the line, and renders system insecurity and thermal overloading. The probabilistic DTR forecasting with high percentile values or confidence levels is thus desired for reliable system operations and analysis. Due to the tight schedule in the day-ahead market, obtaining day-ahead DTR forecasts in a fast way (i.e., within 10 minutes) is preferred. The goal of this Chapter is therefore to provide reliable day-ahead hourly probabilistic DTR forecasts with appropriate percentiles in a fast way for transmission safe operations.

Employing DTR forecasting, however, is difficult. The thermal model considering precipitation to calculate the DTR is associated with many weather factors. Impact of these factors on the thermal capacity is complicated to determine. In addition, in view of the changes of weather conditions along a line, weather measurements and predictions are typically point values and may not be available at selected span locations. Especially when considering multiple span locations, making full use of the weather data for probabilistic DTR forecasting is difficult. Moreover, when determining the DTR considering both weather measurements from multiple locations and weather uncertainties, the DTR becomes the minimum of a few distributions. Considering the tight schedule in day-ahead market, many variables

involved in the complex span thermal capacity function, and the large number of selected spans, extract the exact distribution of DTR in a fast way is difficult.

In this Chapter, a novel approach for probabilistic DTR forecasting is developed to overcome all the above difficulties. Section 4.2 reviews the thermal models used to calculate the conductor thermal capacity, weather data availability and usage, existing approaches for deterministic and probabilistic DTR forecasting, and model validation. In Section 4.3, impact analysis of the span thermal capacity with respect to relevant weather factors is conducted to identify major factors. Insights and implications are discussed on selecting critical spans and installing DTR monitoring systems. To utilize all available data sources and to improve available weather predictions from nearby weather stations, a spatio-temporal regression model with well-selected explanatory variables is established to provide weather predictions and variances.

In Section 4.4, a probabilistic modeling of span thermal capacity and line thermal rating is developed to maintain model accuracy while reducing computational costs. Since the DTR is calculated as the minimum of span thermal capacities, quick and proper modeling of the span thermal capacity is required. Span thermal capacity is directly modeled with a pre-assumed distribution and mean and variance approximately derived from the span thermal capacity function through Taylor series expansion. In view of the difficulty of taking derivative of the max function involved in the thermal equation, the max function is rewritten in form of absolute value function. The cumulative distribution function (CDF) of DTR is then calculated analytically by taking the minimum of a few given distributions. Appropriate percentile of DTR is provided as a high and safe limit for transmission operators.

In Section 4.5, numerical testing results of two examples are presented. In Example 1, a short transmission line segment is used to evaluate the thermal model, the regression weather model, the assumption of truncated Gaussian for span thermal capacity, the safe rating and computational efficiency. In Example 2, a long line is investigated. Thermal capacities at six selected spans are forecast. The CDF

and percentiles of the final DTR are provided and compared with STR and the actual DTR to demonstrate the efficiency of our method.

4.2 Literature Review

This section reviews recent studies on DTR. Subsection 4.2.1 reviews the thermal models investigated to calculate the conductor thermal capacity. Subsection 4.2.2 reviews the availability and the usage of weather data. Subsection 4.2.3 reviews probabilistic DTR modeling considering weather uncertainties and multiple span locations. Subsection 4.2.4 reviews the validation of DTR models.

4.2.1 Heat Balance Thermal Model

The thermal capacity of a conductor is calculated using the heat balance thermal model, which has been investigated in the *IEEE Standard* [6] and *CIGRE Standard* [7], assuming that the sum of heat gains equals the sum of heat losses of a conductor:

$$I^2 \cdot R\left(T_c\right) + Q_s = Q_r + Q_c + M \cdot C_p \cdot \frac{dT_c(t)}{dt} . \quad (4.1)$$

In (4.1), $I^2 \cdot R(T_c)$·is the joule heat from the current, and resistance R is a function of conductor temperature T_c. Solar radiation heat gain Q_s is calculated based on solar heat density Q_{se} which is roughly calculated using polynomial fittings with a specified air condition. Conductor radiation heat loss Q_r is determined by T_c and ambient temperature T_a. Convection cooling Q_c is a function of T_c, T_a, wind speed V, and wind angle ϕ (the angle between the conductor axis direction and direction of wind). Term $M \cdot C_p$ is the heat capacity of the conductor.

Both *IEEE* and *CIGRE Standards* did not consider precipitation for simplification. The thermal model used in [8] and [9] took into account the cooling effects from precipitation and the subsequent evaporating from the conductor. To calculate the thermal capacity at a selected span location, the heat-balance equation in [8], which considers cooling effect Q_e from precipitation, is used:

$$I^2 \cdot R(T_c) + Q_s = Q_r + Q_c + Q_e + M \cdot C_p \cdot \frac{dT_c}{dt}, \qquad (4.2)$$

where

$$Q_e = m \cdot \left[L_e(T_e) + c_w \cdot (T_e - T_a) \right] = f_e(P_r, RH, P_a). \qquad (4.3)$$

In (4.3), m is the mass flus evaporating from the conductor surface, L_e is the specific latent heat of evaporation of water, T_e is the evaporation temperature, and c_w is the specific heat capacity of liquid water. Details can be found in (13)-(25) in [8]. Q_e is a function of precipitation rate P_r, relative humidity RH, and air pressure P_a. Equation (4.2) is now related to weather factors (Q_{se}, T_a, V, ϕ, P_r, RH, P_a) and conductor temperature T_c as:

$$I^2 \cdot R(T_c) + Q_s(Q_{se}) = Q_r(T_c, T_a) + Q_c(T_c, T_a, V, \phi) + Q_e(P_r, RH, P_a) + M \cdot C_p \cdot \frac{dT_c}{dt}, \qquad (4.4)$$

The normal thermal capacity is calculated under steady condition with the assumption that T_c reaches equilibrium and the derivation dT_c/dt becomes zero. With measurements of seven weather factors (Q_{se}, T_a, V, ϕ, P_r, RH, P_a) given at a selected span location and the maximum conductor temperature T_c^{max}, the maximum permissible current I_{max} as the span normal thermal capacity can be calculated as:

$$I_{\max} = \sqrt{\frac{Q_r(T_c^{\max}, T_a) + Q_c(T_c^{\max}, T_a, V, \phi) + Q_e(P_r, RH, P_a) - Q_s(Q_{se})}{R(T_c^{\max})}}. \qquad (4.5)$$

This thermal model was validated in [9] based on weather data from an on-site weather station and current sensors installed on a live transmission line. However, the enhanced model involves many weather factors and complicates the span thermal capability function. Impacts of these factors on the thermal capacity need to be determined.

4.2.2 Weather Factors and Data Availability

To calculate the DTR of a transmission line, measurements or predictions of relevant weather factors at selected span locations are required. Weather factors involved in (5) include ambient temperature, wind speed, wind angle, solar radiation, precipitation rate, relative humidity, and air pressure. To implement an effective DTR system, it is necessary to install weather sensors on transmission line towers with a tradeoff between accurate estimations and equipment investments. In certain DTR systems, weather sensors are installed on selected line towers as shown in Figure 4.2, and typically measure ambient temperature and wind [10], [11]. A few advanced sensors and devices could provide predictions of temperature and wind [12]. Locations of sensor installation are typically selected in the DTR system design phase. In [13], critical spans were roughly selected based on prevailing wind direction, transmission line layout, and geographical information. In [14], a heuristic approach was developed for identifying the number and locations of critical spans based on the use of historical simulated weather.

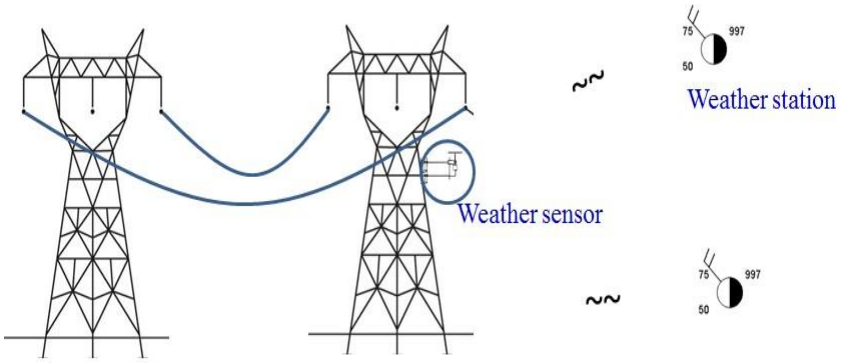


Figure 4.2. Weather data sources of DTR systems.

If no weather sensors are installed on a line, data from weather services such as *Weather Underground* and *National Numerical Weather Predictions* (NWP) are collected [15]-[17]. These weather services generally provide measurements and predictions of nearby temperature, wind, precipitation and relative humidity collected from nearby weather stations as shown in Figure 4.2. Since solar heat density

Q_{se} is generally not available from nearby weather stations, it is modeled as a polynomial function of calendar date, hour index, and air condition as in [6].

4.2.3 DTR with Multiple Spans and Weather Uncertainties

When considering multiple span locations, if no weather are available at pre-selected span locations in the design phase, weather data obtained at nearby locations were interpolated into these span locations to rate an entire line [18]. Point thermal capacity for each span is calculated and the minimum is chosen as the rating of an entire line. Deterministic DTR approaches do not consider the inherent uncertainties in weather. To handle weather uncertainties, a few probabilistic and fuzzy approaches were developed. In [19], Monte Carlo Sampling (MCS) method based on probability density functions (PDFs) of weather factors and the heat balance thermal model was used. In [20], fuzzy numbers were used to describe the uncertainties of input weather data. The thermal capacity of each line segment was then computed using fuzzy calculations. The MCS and fuzzy-based approaches require either appropriate distributions or the membership functions of each variable involved. When considering both weather uncertainties and multiple spans, the above methods require generating individual distribution or membership function for each weather factor, span thermal capacity and the final line rating, and thus become computationally complex.

4.2.4 Model Validation

To validate DTR models, ideally real transmission measurements such as line loading and conductor temperatures are required. A few papers summarized DTR projects with weather sensors measuring conductor temperatures or Phasor Measurement Unit (PMU) measuring line loading [9], [19], [21]. The real line loading or conductor temperature is measured and compared with the maximum current capability or conductor temperature to ensure the measured values do not exceed the designed maximum. Some papers conducted simulations based on *IEEE* test systems with geographical features and nearby weather data to verify the secure line loading and the benefits of DTR implementation [13], [22]. Some

studies lacked of full model validation since they only compared the estimated DTR with the calculated DTR [20], [23]. Considering that the thermal model is widely accepted and the calculated DTR could be treated as reliable; comparing with the calculated DTR is viewed as the state-of-the-practice in industry.

4.3 Impacts of Weather Factors and Weather Modeling

This section presents the impact analysis of relevant weather on the span thermal capacity and the modeling of weather. Subsection 4.3.1 analyzes the span thermal capacity with respect to all weather factors to identify the major ones. Subsection 4.3.2 presents a spatio-temporal regression model to provide weather predictions based on all available data sources.

4.3.1 Impact Analysis and Implications

As introduced in Section 4.1, overhead transmission line could be tens of kilometers long and consist of hundreds of spans in multiple segments. Thus, span thermal capacities change as the changes of locations and the varying weather along the line. As reviewed in Section 4.2.1, the enhanced thermal model considering precipitation is associated with seven weather factors. Changes of span thermal capacities with these seven weather factors are shown in Figure 4.3. The span thermal capacity changes drastically with respect to wind and temperature. To identify the most influential weather factors, impact levels of weather factors are analyzed. The impact level is defined as relative percentage changes of the maximum capacity to the minimum capacity as the weather factor changes:

$$Impact\ Level = \left[\frac{\text{max capacity - min capacity}}{\text{min capacity}}\right]\times 100\%. \qquad (4.6)$$

Note that when analyzing one weather factor, the rest are fixed at the basis values commonly used in STR as specified in Table 4.1. Other ambient parameters and conductor characteristic follow settings in Example 4.1 in [6].

Among all weather factors in (5), wind speed is reported as the most influential one. Therefore, impact levels of other weather factors with five different wind speeds (i.e. 0.6 m/s, 2 m/s, 4 m/s, 8 m/s, and 12 m/s) are calculated and presented in Table 4.2. Based on the impact levels, wind direction, ambient temperature, solar radiation, and precipitation rate are the most influential factors to span thermal capacities. Non-influential weather factors *RH* and P_a can be fixed at the basis values and uncertainties of these factors will not be considered in DTR. Moreover, the effect of wind angle is significant especially when wind speed is high. This indicates when selecting span locations for sensor installation, line axis direction and terrain information should be considered. Since the design of DTR monitoring system is not the focus of this Chapter, we assume weather sensors are installed on certain pre-selected critical spans and our goal is to develop DTR modeling with the given design.

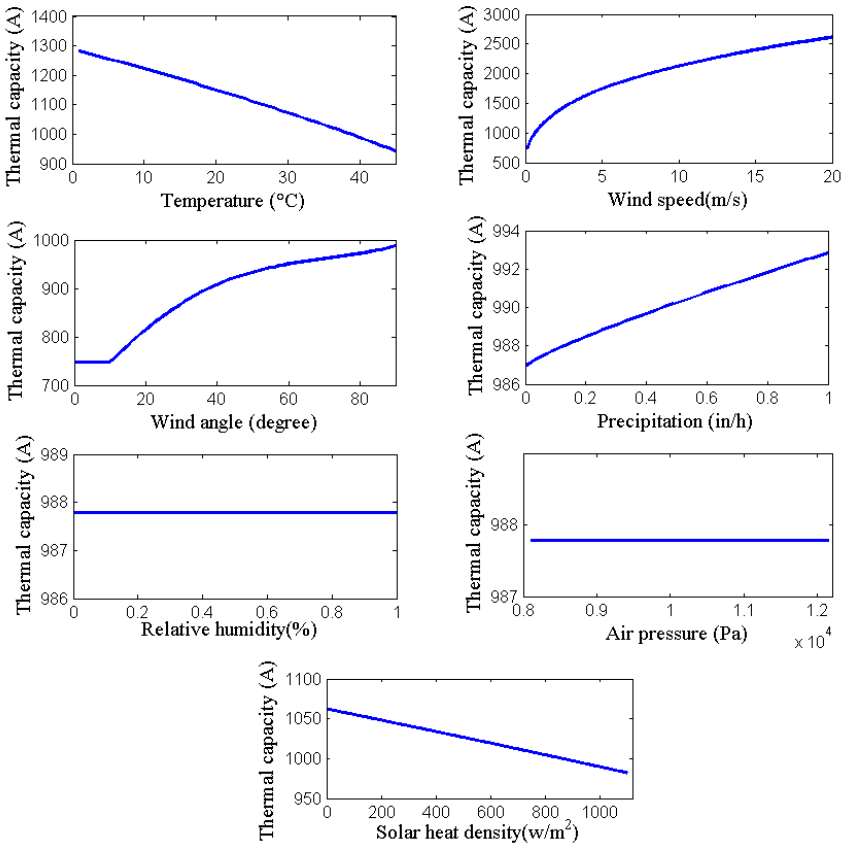


Figure 4.3. Span thermal capacities with changes in weather factors.

TABLE 4.1 BASIS VALUES OF WEATHER FACTORS

Maximum conductor temperature	T_c^{max}	100°C
Ambient temperature	T_a	40°C
Wind speed	V	0.6m/s
Wind angle	ϕ	90
Solar heat density	Q_{se}	1000w/m^2
Precipitation rate	P_r	0.1in/h
Relative humidity	RH	50%
Air pressure	P_a	101.325kPa

TABLE 4.2 THERMAL CAPACITY IMPACT LEVELS (%) WITH WEATHER FACTORS

Weather factors	Range	Wind speed V (m/s)				
		0.6	2	4	8	12
T_a	0–45°C	36.21	35.22	34.83	34.54	34.40
ϕ	0–90°	31.92	53.09	55.17	56.56	57.12
Q_{se}	0–1100 w/m^2	8.16	4.52	3.05	2.05	1.62
P_r	0–1 in/h	0.60	1.09	1.46	1.95	2.31
RH	0–100%	0	0	0	0	0
P_a	80–120kPa	0	0	0	0	0

4.3.2 Spatio-Temporal Regression Model

As reviewed in Section 4.2.2 on the weather data sources and availability, it is important to utilize all available information to obtain appropriate weather predictions to feed into the thermal model. A spatio-temporal regression model is established to provide predictions based on all available weather data as expressed below:

$$w_{k,m}(t) = \beta_{k,m,0} + \sum_{n_1=1}^{3} \beta_{k,m,n_1} w_{k,m,n_1}(t-24) + \beta_{k,m,4} \hat{w}_{k,m}(t) + \sum_{n_2=1}^{4} \beta_{k,m,4+n_2} s_{n_2} + \varepsilon_{k,m}(t), \forall k, m, \quad (4.7)$$

where the index m represents the weather factors including temperature, adjusted wind speed, wind angle, and precipitation rate to be considered, and the index k represent the span location selected. To determine the weather factor $w_k(t)$ at hour t at a targeted span k, lagged measurements $w_k(t\text{-}24)$ from the closest three weather stations (including the targeted location if available) are considered as part of explanatory variables in the regression model. Wind speeds measured from a nearby weather station are generally at a

different height from that of transmission towers. Measurements at height h_0 of a weather station are thus adjusted following the wind profile power law [24] to represent the wind at height h as the span or tower height:

$$V_h = V_{h_0} \cdot \left(h / h_0\right)^a, \qquad (4.8)$$

where the exponent a for different ground types can be found in [24]. To be precise, in this Chapter, parameter a is fitted by historical weather measurements at known heights.

The variable $\hat{w}_{k,m}$ in (4.7) represents the given weather prediction at a targeted location. The use of this explanatory variable would help improve the modeling accuracy. If the value is not available from a DTR monitoring device, interpolation from nearby NWPs by the inverse squared distance technique [25] is used as:

$$w_{k,m} = \frac{\sum_{i=1}^{N}\left(1/d_{n,k}{}^2\right)\cdot w_{n,m}}{\sum_{i=1}^{N}\left(1/d_{nk}{}^2\right)}, \quad \forall k,m, \qquad (4.9)$$

where n is the number of weather stations, $d_{n,k}$ is the distance between the nth nearby weather station and the kth span location.

To better capture the spatial features and achieve predictions at locations where we have no measurements, a set of geographic covariates s_{n_2} including latitude, longitude, distance to valley, and distance to coast are used. The selection of these spatial covariates is based on cross validation. The weather model is validated through historical measurements and predictions provided by National Renewable Energy Laboratory weather stations and Weather Underground.

4.4 Probabilistic DTR Modeling

A probabilistic model is developed to maintain model accuracy while reducing computational complexity. The DTR is determined as the minimum of selected span thermal capacities, which are modeled as truncated Gaussian variables. Subsection 4.4.1 introduces an approximation approach to model span thermal capacities. Subsection 4.4.2 describes a fast way to obtain the distribution and appropriate percentiles of the DTR.

4.4.1 Probabilistic Span Thermal Capacity Modeling

Span thermal capacity is a nonlinear function of multiple weather factor variables derived from the heat-balance equation as shown in (4.5) in 4.2.1. It is difficult to extract the exact distribution. Traditional approaches such as MCS require appropriate distributions of weather factors and are computationally complex. Since weather factors are generally within ranges and assumed normally distributed, distribution of the span thermal capacity is approximately treated as truncated normal, which is a probability distribution of a normally distributed random variable with bounded values. Mean and variance of each span thermal capacity are derived through the heat-balance equation with means and variances of weather factors obtained from the spatio-temporal regression model. Truncated normal distribution has been used in [26] and [27] to model line thermal capacities in view of the cut-off feature of weather factors and thermal capacities.

Truncated normal distribution of the thermal capability I_k at span k is thus denoted as $N(\mu_k, \sigma_k^2; a_k, b_k)$, where a_k and b_k are determined as extreme cases of I_k. For example, upper limit is based on the lowest temperature, largest wind speed, precipitation rate, and perpendicular wind direction. The probability density function (PDF) and cumulative distribution function (CDF) of the truncated normal distribution can be referred to subsections 3.1 and 3.3 in [28]. Approximations for mean and variance of I_k are thus achieved using Taylor series expansion of function f:

$$I_k = \sqrt{\frac{Q_r\left(T_c^{\max}, T_{a,k}\right) + Q_c\left(T_c^{\max}, T_{a,k}, V_k, \phi_k\right) + Q_e\left(P_{r,k}\right) - Q_s}{R\left(T_c^{\max}\right)}} = f(w_k) = f\left(T_{a,k}, V_k, \phi_k, P_{r,k}\right). \quad (4.10)$$

at the expansion point $\theta_k = \left(\mu_{T_a,k}, \mu_{V,k}, \mu_{\varphi,k}, \mu_{P_r,k}\right)$ chosen as means of weather factors. In this case, the first order Taylor series approximation of *f* becomes

$$I_k = f\left(\theta_k\right) + \sum_{m=1}^{4} f'_{w_{m,k}}\left(w_{m,k} - \mu_{m,k}\right) + R_k, \quad (4.11)$$

where R_k is the remaining term.

The approximation for $E(I_k)$ is written as:

$$\begin{aligned} \mathrm{E}\left(I_k\right) &= \mathrm{E}\left[f\left(\theta_k\right) + \sum_{m=1}^{4} f'_{w_{m,k}}\left(\theta_k\right) \cdot \left(w_{m,k} - \mu_{m,k}\right) + R_k \right] \\ &\approx \mathrm{E}\left[f\left(\theta_k\right)\right] + \sum_{m=1}^{4} \mathrm{E}\left[f'_{w_{m,k}}\left(\theta_k\right) \cdot \left(w_{m,k} - \mu_{m,k}\right)\right] \\ &= \mathrm{E}\left[f\left(\theta_k\right)\right] + \sum_{m=1}^{4} f'_{w_{m,k}}\left(\theta_k\right) \cdot \mathrm{E}\left[\left(w_{m,k} - \mu_{m,k}\right)\right] \\ &= \mathrm{E}\left[f\left(\theta_k\right)\right] = f\left(\theta_k\right), \end{aligned} \quad (4.12)$$

and the variance of I_k is calculated as:

$$\mathrm{Var}\left(I_k\right) = \mathrm{E}\left\{\left[f\left(w_k\right) - \mathrm{E}\left(f\left(w_k\right)\right)\right]^2\right\} \approx \mathrm{E}\left\{\left[f\left(w_k\right) - f\left(\theta_k\right)\right]^2\right\}. \quad (4.13)$$

Expand *f* with the first order Taylor series approximation, and the variance becomes:

$$\begin{aligned} \mathrm{Var}(I_k) &\approx \mathrm{E}\left\{\left[\sum_{m=1}^{4} f'_{w_{m,k}}\left(\theta_k\right) \cdot \left(w_{m,k} - \mu_{m,k}\right)\right]^2\right\} \\ &= E\left\{\sum_{m=1}^{4} f'^2_{w_{m,k}}\left(\theta_k\right)\left(w_{m,k} - \mu_{m,k}\right)^2 + 2\sum_{i=1}^{4}\sum_{\substack{j=1 \\ j\neq i}}^{4} f'_{w,i}\left(\theta_k\right) f'_{w_j}\left(\theta_k\right)\left(w_{i,k} - \mu_{i,k}\right)\left(w_{j,k} - \mu_{j,k}\right)\right\} \\ &= \sum_{m=1}^{4} f'^2_{w_{m,k}}\left(\theta_k\right)\mathrm{Var}\left(w_{m,k}\right) + 2\sum_{i=1}^{4}\sum_{\substack{j=1 \\ j\neq i}}^{4} f'_{w_{k,i}}\left(\theta_k\right) f'_{w_{j,k}}\left(\theta_k\right)\mathrm{Cov}\left(w_{i,k}, w_{j,k}\right), \end{aligned} \quad (4.14)$$

where Var($w_{m,k}$) is the variance of weather factor m for critical span k obtained from the spatio-temporal regression model. Since weather factors can be treated as independent random variables, the covariance of each two weather factors $\mathrm{Cov}(w_{i,k}, w_{j,k})$ is zero.

When taking the derivative of function f as expressed in (4.10), the calculation of convection cooling Q_c involves a max function [6] as

$$Q_C = \max(Q_{c1}, Q_{c2}) . \qquad (4.15)$$

In view of the difficulty of taking derivative of the max function involved in the thermal equation, the max function is rewritten in form of the absolute value function as:

$$Q_C = \max(Q_{c1}, Q_{c2}) = \frac{Q_{c1} + Q_{c2} + |Q_{c1} - Q_{c2}|}{2} . \qquad (4.16)$$

The first derivate of Q_c based on (4.17) are obtained as:

$$Q'_C = \frac{Q'_{c1} + Q'_{c2} + (Q'_{c1} - Q'_{c2}) sign(Q_{c1} - Q_{c2})}{2} . \qquad (4.17)$$

Given (4.17), the entire derivative $f'_{w_{m,k}}$ can thus be calculated.

4.4.2 Dynamic Thermal Rating Modeling

When considering multiple span locations, rating of a line is determined by the most critical span thermal capacities, and DTR is written as the minimum of selected span thermal capacity random variables:

$$Y = \min_{1 \le k \le K} I_k , \qquad (4.18)$$

where I_k is the thermal capacity at span k, Y is the line thermal rating, and K is the number of spans considered. The cumulative distribution function (CDF) F_Y of Y is expressed as:

$$
\begin{aligned}
F_Y(y) &= P\{\min(I_1, I_2, ..., I_K) \le y\} = 1 - P\{\mathrm{Y} > y\} \\
&= 1 - P\{I_1 > y, I_2 > y, ..., I_K > y\} \\
&= \sum_{k=1}^{K} F_{I_1}(y) - F_{I_1 I_2}(y, y) - F_{I_1 I_3}(y, y) - ... - F_{I_{K-1} I_K}(y, y) + F_{I_1 I_2 I_3}(y, y, y) + ... \\
&\quad + F_{I_{K-2} I_{K-1} I_K}(y, y, y) + ... + (-1)^{K-1} F_{I_1 I_2 ... I_K}(y, y, ..., y).
\end{aligned} \tag{4.19}
$$

In (4.19), $F_{I_k}(y)$ is the CDF of the truncated normal distribution expressed as:

$$
F_{I_k}(y) = F_{I_k}(y; \mu_k, \sigma_k^2, a_k, b_k) = \frac{\Phi(\mu_k, \sigma_k^2; y) - \Phi(\mu_k, \sigma_k^2; a_k)}{\Phi(\mu_k, \sigma_k^2; b_k) - \Phi(\mu_k, \sigma_k^2; a_k)}, \tag{4.20}
$$

where Φ is the CDF of standard normal distribution. The rest terms in (4.19) are CDFs of the joint distributions of truncated normal variables. To reduce computational complexity, only correlations between consecutive span thermal capacities are considered in determining the joint distributions. Therefore, we have

$$
F_{I_{k-1} I_k}(y_{k-1}, y_k) = \int_{a_{k-1}}^{b_{k-1}} \int_{a_k}^{b_k} f_{I_{k-1} I_k}(y_{k-1}, y_k) dy_{k-1} dy_k , \tag{4.21}
$$

and

$$
F_{I_{k-2} I_{k-1} I_k}(y, y, y) = F_{I_{k-2} I_{k-1}}(y, y) \cdot F_{I_k}(y) . \tag{4.22}
$$

As the individual and joint distributions are obtained from (4.18)-(4.22), the CDF of the DTR can thus be calculated. The *p*th percentile of DTR y_p is then obtained by

$$
\frac{p}{100} = P\left(X \le x_p\right) = P\left(-\infty < X \le x_p\right) = \int_{-\infty}^{x_p} f_x(x) dx = F_x(x_p) . \tag{4.23}
$$

Low percentiles such as 2nd, 1st, and 0.1th of the line thermal rating as safe limits are the most important parameters for secure transmission system operations.

4.5 Numerical Testing

Numerical testing on one short line segments and a long transmission line are presented in this Section. Results demonstrate the performance of weather modeling and the improvements of DTR compared with conventional approaches. Truncated normal distribution assumption is verified, and the computational efficiency is verified through comparisons with Monte Carlo Simulation. In all testing, characteristics of conductors follow the settings in [15]: normal T_c^{Max} = 100ºC, conductor diameter = 28.1 mm, emissivity on conductor's surface = solar absorptivity of the conductor = 0.5 and clear air atmosphere. Thermal capacities used in numerical testing are expressed as below:

STR: A constant value calculated with T_a = 40ºC, V = 0.6m/s, ϕ = 90, and equals to 945A.

Actual span thermal capacity: a value calculated with actual weather measurements at a span.

Actual DTR: a value calculated as the minimum of all calculated span thermal capacities.

Estimated deterministic span thermal capacity: a value calculated with predicted means of weather factors.

Estimated deterministic DTR: a value as the minimum of deterministic estimated span thermal capacities.

Estimated probabilistic span thermal capacity: a distribution calculated with predicted means and variances of weather.

Estimated probabilistic span thermal capacity: the minimum of probabilistic estimated span thermal capacities.

4.5.1 Example 1

Numerical testing results on a short transmission line segment and a long line are presented in this Section. Characteristics of conductors follow the settings in [6] as normal T_c^{Max} = 100°C, conductor

diameter = 28.1mm, emissivity on conductor's surface = solar absorptivity of the conductor = 0.5, and clear air atmosphere. In the following testing, the STR is a constant value calculated as 945A. Actual span thermal capacities and the line rating are values calculated with actual interpolated weather measurements.

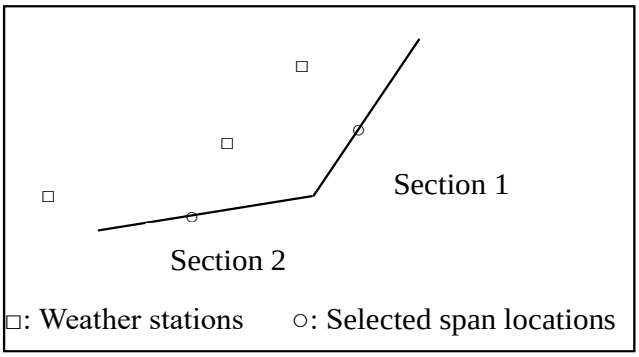


Figure 4.4. Layout of the transmission line segment and weather stations.

Example 1: A short line segment of a 230-kV transmission line from Lookout Substation to Ridge Plant in Boulder, CO is investigated. The line segment consists of two line sections and twenty spans as depicted in the layout in Figure 4.4. Since there are no weather sensors installed on the transmission towers, weather data of three nearby weather stations from *Weather Underground* are collected and interpolated to selected two span locations. Historical data from January to September in 2014 are used for training of the weather regression model, and October 2014 is used for prediction. Four cases are conducted to demonstrate the efficiency of our method. Case 1 shows the significance of considering cooling effects from precipitation. Case 2 shows the performance of the developed weather model. Case 3 verifies the assumption of the truncated Gaussian for the span thermal capacity and compares with the Monte Carlo Sampling. Case 4 demonstrates the improvements of considering multiple spans and determining DTR probabilistically.

Case 1: Impact of cooling effects from precipitation and the subsequent evaporating from the conductor on conductor thermal capability is demonstrated. Table 4.3 shows the increased thermal

capacities (%) under different levels of rain rates from considering precipitation. As can be seen, when the rain is heavy, the thermal capacity could improve over 5%.

Table 4.3. Increased thermal capacity from considering precipitation.

Conditions with rain rate *Pr* (mm/h)		***Pr***	**Increased thermal capacity (%)**
Very light rain	$Pr < 0.25$	0.2	0.27
Light rain	$0.25 \leq Pr < 1$	0.5	0.58
Moderate rain	$1 \leq Pr < 4$	3	2.61
Heavy rain	$4 \leq Pr < 16$	10	7.01
Very heavy rain	$16 \leq Pr < 50$	30	16.94
Extreme rain	$Pr \geq 50$	50	25.01

Case 2: This case demonstrates the performance of the regression model to provide weather predictions at selected spans. Weather predictions obtained from the regression model are compared with the interpolated measurements at span locations. As shown in Figure 4.5, prediction results from the regression model provide a general trend for each weather factor. With the standard derivations obtained from the regression model, the shadow areas in Figure 4.5 illustrate the two-sigma ranges for individual weather factors.

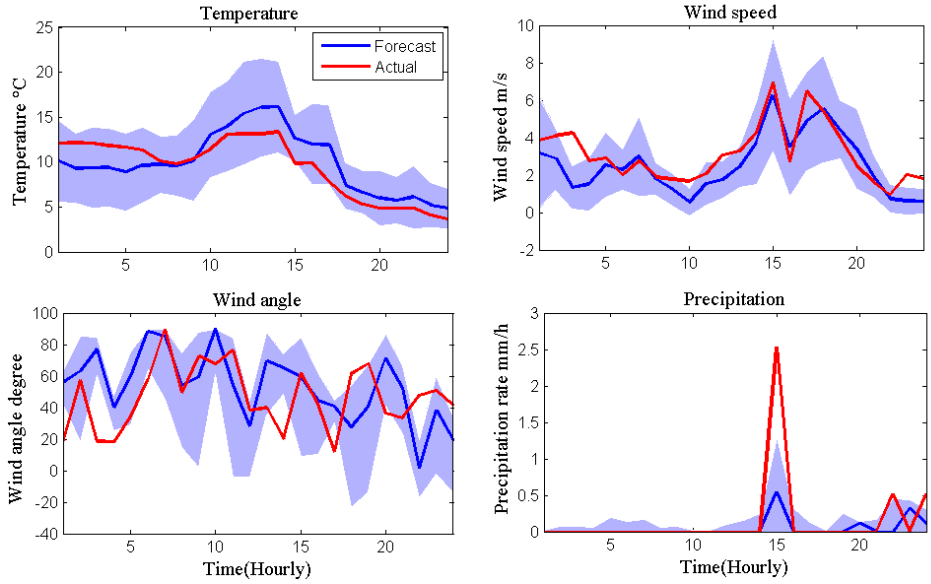


Figure 4.5. Weather predictions at span 1 on Oct. 17th, 2014.

Case 3: This case verifies the assumption of the truncated Gaussian distribution for the span thermal capacity. The MCS is conducted to obtain the distribution of the span thermal capacity based on the following distribution assumption for weather factors: normal for T_a, ϕ, P_r, and Weibull distribution for V. Same means and variances of each weather factor are used for both our method and the MCS. Quantile-Quantile (Q-Q) plots of span thermal capacities obtained from our approach verse MCS for spans 1 and 2 are depicted in Figure 4.6. As can be seen, distribution of span thermal capacities obtained from our approach has a nice fit with that of the MCS. Moreover, expected values and important percentiles (0.1th, 1st, and 2nd) of the DTR obtained from our approach and the MCS are compared as shown in Figure 4.7. The computational time to obtain DTR is summarized in Table 4.4. In general, performance of the two methods is similar; however, our approach is nearly five times faster than the MCS.

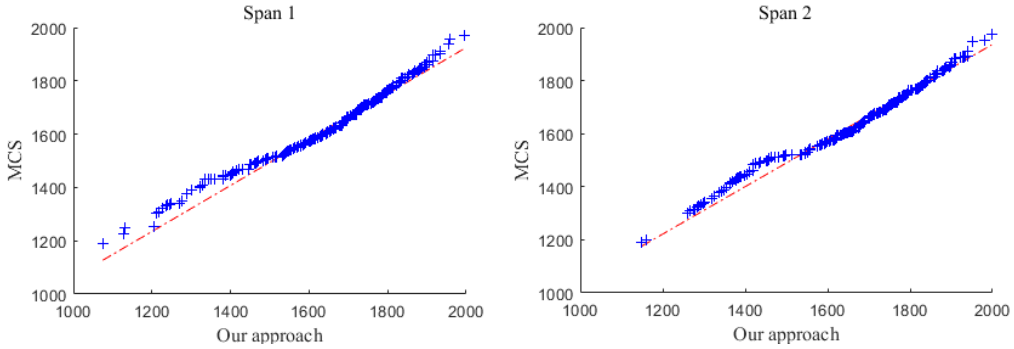


Figure 4.6. Quantile-Quantile plot of the span thermal capacity obtained from our approach versus MCS.

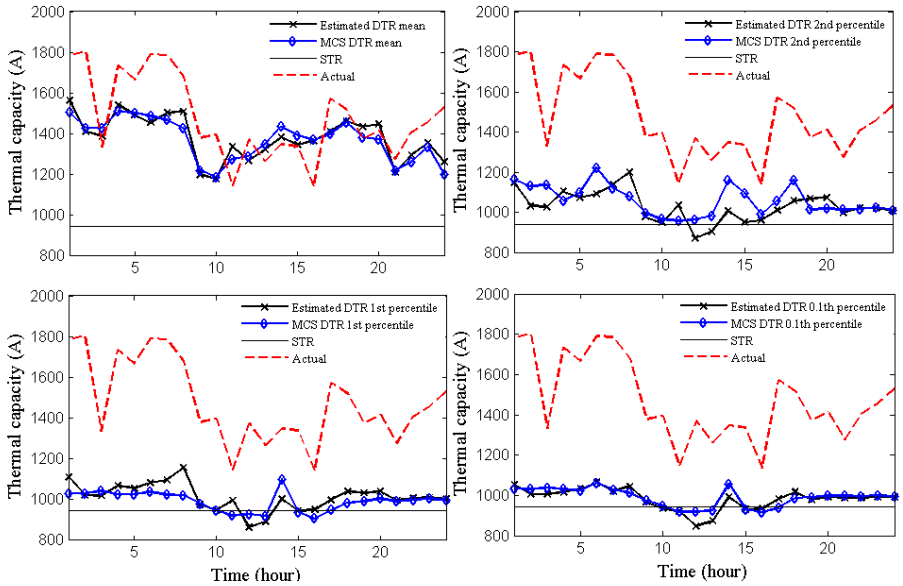


Figure 4.7. Comparisons of estimated DTR by our approach and MCS.

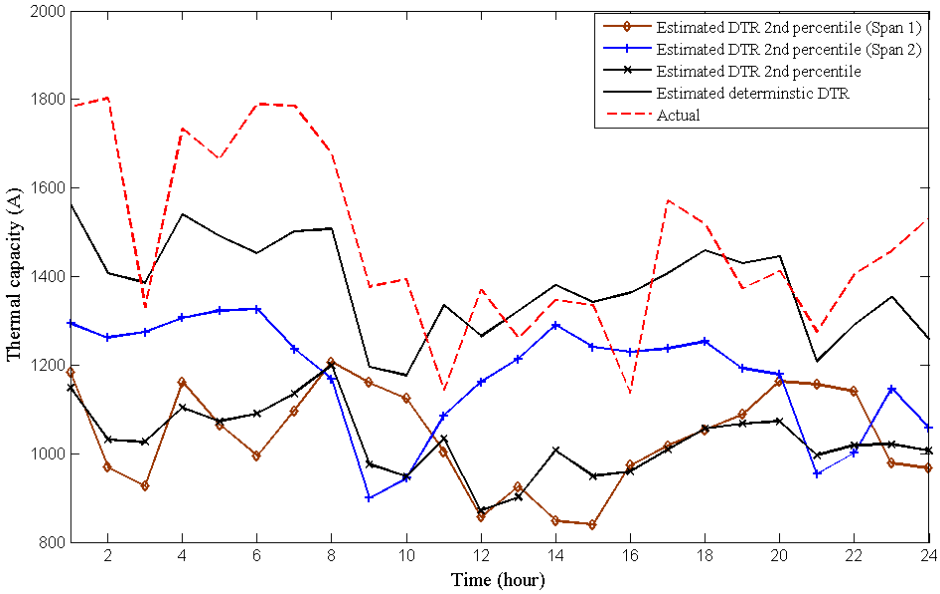


Figure 4.8. Comparisons of 2nd percentile estimated DTRs.

Case 4: This case demonstrates the improvements of our probabilistic DTR forecasts compared to DTR forecasts considering a single span location and deterministic DTR forecast. The 2nd percentile of the estimated DTR from our approach with consideration of weather conditions at both spans is compared with the estimated DTRs determined at a single span location. As results depicted in Figure 4.8, the estimated DTR based on a single span location could result in large bias since the measurement location

may not be at a critical span location. Moreover, the estimated probabilistic DTR is compared with the deterministic one. Variances of weather predictions from actual measurements would lead to large variances in thermal capacities, especially for wind that is difficult to predict with a high accuracy. As is shown, the estimated deterministic DTR is likely to exceed the actual DTR for some time instances due to inherent weather uncertainties. However, probabilistic DTR at percentiles below 2nd do not exceed the actual DTR and thus provides high rating without scarifying system security. Therefore, we suggest using probabilistic DTR at low percentiles to ensure secure system operations.

4.5.2 Example 2

A long 230-kV transmission line between Lookout Substation and Plan End Power Plant in Colorado, United States identified in Google Earth is investigated for the study. This line contains six sections and totally sixty-eight spans. The prevailing wind is West-East direction in this area. As shown in Figure 4.9, one span is selected on each line section as the critical span for DTR modeling. Among them, S1, S3 and S5 are in a North-South line direction while S2, S4, and S6 are in West-East direction. To forecast the DTR, historical weather measurements and weather predictions from twelve nearby weather stations provided by *Weather Underground* are collected and interpolated into the selected six spans. DTRs on October 17 and 18 in 2014 are forecast and analyzed below. The computational time of our method to obtain the DTR is summarized and compared with MCS as in Table 4.5.

Figure 4.9. Layout of a 230-kV transmission line in Colorado, US.
[Source: 39°47'33.20" N 10°59.41" W. **Google Earth**. May 31, 2018. November 05. 2019.]

Table 4.5. Computational times (s) of obtaining DTR forecasts in Example 2.

Our Approach	**MCS**
408	2,136

Means of the estimated span thermal capacities at six selected spans are represented in Figure 4.10. Differences between the estimated deterministic DTRs and the actual DTRs are mainly from the weather forecasting errors. The different span thermal capacities demonstrate that the DTR modeling based on a single span location is insufficient. Important percentiles (2nd, 1st, and 0.01th) of hourly DTR forecasts are provided in Figure 4.11. With 2% and lower percentiles, the estimated DTRs do not exceed the actual DTR and thus provides high rating without scarifying system security. Our method demonstrates the increased line thermal capacity from using 2nd percentile of the DTR while maintaining secure system operations.

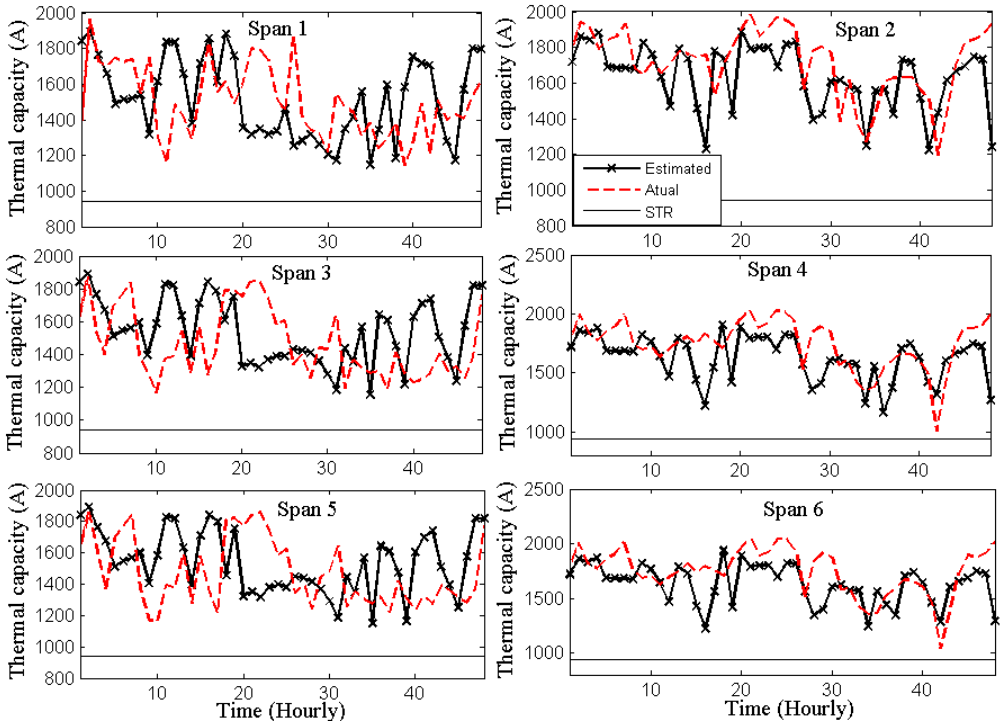


Figure 4.10. Estimated, actual, and static span thermal capacities.

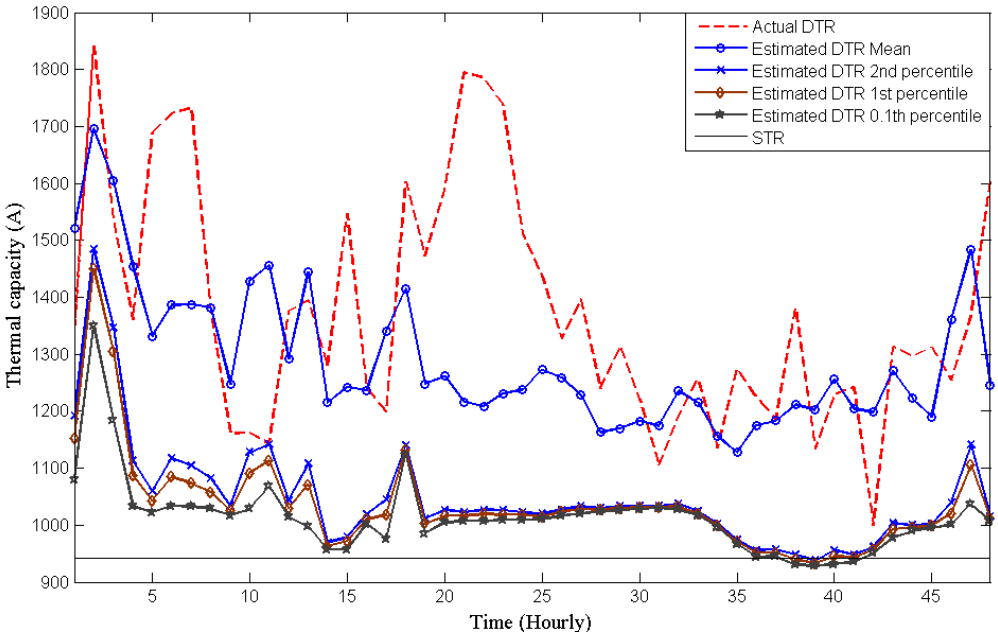


Figure 4.11. Hourly DTR forecasts (means, 0.1th, 1st and 2nd percentiles) on Oct. 17-18, 2014.

4.6 Conclusion

In this Chapter, a novel probabilistic DTR forecasting approach is developed. Both the topology of transmission lines and the inherent weather uncertainties are considered. Distribution of the estimated DTR and important percentile values are obtained to ensure secure and reliable operations. The work implies a promising probabilistic approach to operate networks in the future.

References

[1] T. O. Seppa, "Reliability and real time transmission line ratings," The Valley Group – A Nexans Company, Ridgefield, CT, Tech. Rep., 2007.

[2] W. Wang and S. Pinter, "Dynamic line thermal rating systems for transmission lines, U.S. Department of Energy," Tech. Rep., April 2014.

[3] D. Douglass, D. C. Lawry, A. Edris and E. C. Bascom, "Dynamic thermal ratings realized circuit load limits," IEEE Computer Applications in Power, vol. 13, no. 1, pp. 38-44, 2000.

[4] T. O. Seppa and A. Salehian, "Random variability and uncertainty of transmission line thermal ratings," in Proc. 2006 IEEE Power and Energy Society General Meeting, July 2006.

[5] E. Cloet and J. L. Lilien, "Uprating transmission lines through the use of an innovative real-time monitoring system," IEEE PES 12th Int. Conf. on Transmission and Distribution Construction, Operation and Live Line Maintenance (ESMO), May 2011.

[6] Institute of Electrical and Electronics Engineers (IEEE), IEEE Standard 738-2012: Standard for calculating the current-temperature relationship of bare overhead conductors, 2012.

[7] CIGRE, "Thermal behavior of overhead conductors," CIGRE 207, August 2002.

[8] P. Pytlak, P. Musilek, E. P. Lozowski, and J. Toth, "Modelling precipitation cooling of overhead conductors," Elect. Power Syst. Res., vol. 81, no. 12, pp. 2147–2154, 2011.

[9] G. Kosec, M. Maksic, and V. Djurica, "Dynamic thermal rating of power lines–Model and measurements in rainy conditions," International Journal of Electrical Power & Energy Systems, vol 91, pp 222-229, October 2017.

[10] B. P. Bhattarai, "Improvement of transmission line ampacity utilization by weather-based dynamic line rating," IEEE Transactions on Power Delivery, vol. 33, no. 4, pp. 1853-1863, Aug. 2018.

[11] C. R. Black and W. A. Chisholm, "Key Considerations for the Selection of Dynamic Thermal Line Rating Systems," IEEE Transactions on Power Delivery, vol. 30, no. 5, pp. 2154-2162, Oct. 2015.

[12] D. L. Alvarez, F. Faria da Silva, E. E. Mombello, C. L. Bak, J. A. Rosero, and D. L. Ólason, "An approach to dynamic line rating state estimation at thermal steady state using direct and indirect measurements," Elect. Power Syst. Res., vol. 163, pp. 599-611, 2018.

[13] Oncor Electricity Delivery Company, "Dynamic line rating Oncor electric delivery smart grid program," Tech. Rep., 2013.

[14] M. Matus, D. Sáez, M. Favley, C. Suazo-Martínez, J. Moya, G. Jiménez-Estévez, R. Palma-Behnke, G. Olguín, and P. Jorquera, "Identification of critical spans for monitoring systems in dynamic thermal rating," IEEE Trans. on Power Delivery, vol. 27, no. 2, 2012.

[15] R. J. Hemparuva, S. P. Simon, S. Kinattingal, and N. P. Padhy, "Geographic information system and weather based dynamic line rating for generation scheduling," Engineering Science and Technology, vol 21, pp. 564–573, 2018.

[16] J. L. Aznarte and N. Siebert, "Dynamic line rating using numerical weather predictions and machine learning: A case study," in IEEE Trans. on Power Delivery, vol. 32, no. 1, pp. 335-343, Feb. 2017.

[17] A. Michiorri, P. C. Taylor and S. C. E. Jupe, "Overhead line real-time rating estimation algorithm: description and validation," Proc. IMechE, vol. 224, 2009.

[18] K. Morozovska and P. Hilber, "Study of the monitoring systems for Dynamic Line Rating," Energy Procedia, vol. 105, pp. 2557-2562, 2017.

[19] D. M. Greenwood et al., "A Comparison of real-time thermal rating systems in the U.S. and the U.K.," IEEE Transactions on Power Delivery, vol. 29, no. 4, pp. 1849-1858, Aug. 2014.

[20] H. Shaker, M. Fotuhi-Firuzabad and F. Aminifar, "Fuzzy dynamic thermal rating of transmission lines," IEEE Transactions on Power Delivery, vol. 27, no. 4, pp. 1885-1892, 2012.

[21] A. Michiorri et al., "Forecasting for dynamic line rating," Renewable Sustain. Energy Rev., vol. 52, pp. 1713–1730, Dec. 2015.

[22] F. Teng, R. Dupin, A. Michiorri, G. Kariniotakis, Y. Chen, and G. Strbac, "Understanding the benefits of dynamic line rating under multiple sources of uncertainty," IEEE Trans. Power Syst., vol. 33, no. 3, pp. 3306– 3314, May 2018.

[23] J. Zhan, C. Y. Chung, and E. Demeter, "Time series modeling for dynamic thermal rating of overhead lines," IEEE Trans. Power Syst., vol. 32, no. 3, pp. 2172–2182, May 2017.

[24] S. A. Hsu, Eric A. Meindl, and David B. Gilhousen, "Determining the power-law wind-profile exponent under near-neutral stability conditions at Sea," J. Appl. Meteor., 33 (6), pp 757–765, 1994.

[25] M. C. Alexiadis, P. S. Dokopoulos and H. S. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," IEEE Transactions on Energy Conversion, vol. 14, no. 3, Sep., 1999.

[26] M. A. Bucher and G. Andersson, "Robust corrective control measures in power systems with dynamic line rating," IEEE Trans. Power Syst., vol. 31, no. 3, pp. 2034–2043, May 2016.

[27] F. Fan, K. Bell and D. Infield, "Probabilistic real-time thermal rating forecasting for overhead lines by conditionally heteroscedastic auto-regressive models," IEEE Transactions on Power Delivery, vol. 32, no. 4, pp. 1881-1890, Aug. 2017.

[28] J. Burkardt, The truncated normal distribution, 2014, [Online]. Available: http://people.sc.fsu.edu/~jburkardt/ presentations/truncated_normal.pdf.